

Single-cell RNA sequencing defines distinct disease subtypes and reveals hypo-responsiveness to interferon in asymptomatic Waldenstrom's Macroglobulinemia

Received: 30 January 2024

Accepted: 16 January 2025

Published online: 10 February 2025

 Check for updates

Romanos Sklavenitis-Pistofidis^{1,2,3,6}, Yoshinobu Konishi^{1,2,3,6}, Daniel Heilpern-Mallory^{1,2}, Ting Wu², Nicholas Tsakmaklis¹, Michelle P. Aranha^{1,2,3}, Zachary R. Hunter^{1,3}, Alaa K. Ali¹, Junko Tsuji², Nicholas J. Haradhvala², Elizabeth D. Lightbody^{1,2,3}, Katherine Towle¹, Laura Hevenor¹, Rizwan Romee^{1,3}, Edward L. Briercheck^{1,3}, Eric L. Smith^{1,3}, Christine-Ivy Liacos⁴, Efstathios Kastritis⁴, Meletios A. Dimopoulos⁴, Steven P. Treon^{1,3}, Gad Getz^{1,3,5,7} & Irene M. Ghobrial^{1,2,3,7} 

Waldenstrom's Macroglobulinemia (WM) is an IgM-secreting bone marrow (BM) lymphoma that is preceded by an asymptomatic state (AWM). To dissect tumor-intrinsic and immune mechanisms of progression, we perform single-cell RNA-sequencing on 294,206 BM tumor and immune cells from 30 patients with AWM/WM, 26 patients with Smoldering Myeloma, and 23 healthy donors. Despite their early stage, patients with AWM present extensive immune dysregulation, including in normal B cells, with disease-specific immune hallmarks. Patient T and NK cells show systemic hypo-responsiveness to interferon, which improves with interferon administration and may represent a therapeutic vulnerability. *MYD88*-mutant tumors show transcriptional heterogeneity, which can be distilled in a molecular classification, including a *DUSP22/CD9*-positive subtype, and progression signatures which differentiate IgM MGUS from overt WM and can help advance WM research and clinical practice.

Waldenstrom's Macroglobulinemia (WM) is a rare non-Hodgkin, IgM-secreting lymphoplasmacytic lymphoma (LPL) of the bone marrow (BM) with an incidence of 3.8 per million persons per year in the United States^{1–3}. It is consistently preceded by two precursor conditions, IgM

monoclonal gammopathy of undetermined significance (IgM MGUS) and smoldering WM (SWM)^{4–6}. According to the definition of the 2nd International Workshop on WM, asymptomatic patients with a monoclonal IgM protein in their peripheral blood (PB) serum and no

¹Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. ²Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, Cambridge, MA, USA. ³Harvard Medical School, Boston, MA, USA. ⁴Department of Clinical Therapeutics, National and Kapodistrian University of Athens, School of Medicine, Athens, Greece. ⁵Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, MA, USA. ⁶These authors contributed equally: Romanos Sklavenitis-Pistofidis, Yoshinobu Konishi. ⁷These authors jointly supervised this work: Gad Getz, Irene M. Ghobrial.

✉ e-mail: irene_ghobrial@dfci.harvard.edu

morphological evidence of BM infiltration by lymphoma are diagnosed with IgM MGUS, while asymptomatic patients with a monoclonal IgM protein and any amount of BM infiltration are diagnosed with SWM⁷. Together, we refer to IgM MGUS and SWM as asymptomatic WM (AWM). Currently, patients with AWM are not treated until they progress to overt, symptomatic WM due to concerns over the short- and long-term complications of treatment^{8,9}. To improve follow-up recommendations and direct investigations of risk-adapted treatment approaches, we need to better understand which patients with AWM are at the greatest risk of disease progression. In the largest retrospective study to date, conducted on 439 patients with AWM, we identified BM infiltration, serum levels of IgM, β 2-microglobulin, and albumin as independent predictors for progression¹⁰. Using these variables, we developed a prognostic model to stratify patients into low, intermediate, and high-risk groups, which was validated in multiple external datasets^{10,11}. However, this as well as other prediction algorithms leverage clinical variables only, highlighting the need for studies that dissect the biology of disease progression and identify tumor-intrinsic and extrinsic factors that can inform patient prognostication and therapy selection^{10,12–15}.

We previously demonstrated that wild-type status for *MYD88*, which is mutated in more than 90% of patients, may be associated with higher risk of progression to WM, while it was also shown to impact time to progression and risk of histological transformation^{10,14,16–18}. Moreover, the presence of activating *CXCR4* mutations, which are seen in approximately 30–40% of patients with WM and are associated with inferior response to certain therapies and poor outcome, was also shown to be associated with higher risk of progression^{3,19–23}. Deletion of the long arm of chromosome 6 (Del6q), which is the most frequent copy number abnormality encountered in patients with WM (~50%), was also shown to be associated with shorter time to progression^{24–26}. Studies utilizing gene expression profiling (GEP) technology to characterize the WM transcriptome along stages of disease progression have been inconclusive in terms of identifying key gene expression programs associated with progression^{27,28}. Even less is known about the role of the BM immune microenvironment in regulating disease progression to overt WM^{29–35}. Changes in the composition and function of T cells, NK cells and monocytes have been reported, with increased cytotoxic T and NK cells, regulatory T cells (Tregs), and non-canonical monocytes, however it is unclear how these may impact disease progression from early stages to overt WM^{32–34}.

In this study, we investigate tumor-intrinsic mechanisms of progression and comprehensively characterize changes in the BM immune microenvironment with disease progression by performing single-cell RNA and B cell receptor (BCR) sequencing (scRNA-seq & scBCR-seq) on BM tumor and immune cells ($n=294,206$) obtained from 30 patients with AWM and WM, 26 patients with Smoldering Multiple Myeloma (SMM), which is a different type of BM plasma cell premalignancy, and 23 healthy donors (HD). Using this approach, we define the immune hallmarks of WM, identify immune biomarkers and therapeutic vulnerabilities, and describe subtypes of disease as well as tumor-intrinsic signatures of progression.

Results

Immune dysregulation is established early in the course of WM

To characterize changes in the BM immune microenvironment of patients with AWM, we performed single-cell RNA sequencing on 28 BM immune cell samples (IgM MGUS, $n=6$; SWM, $n=19$; WM, $n=3$) obtained from 27 patients, including one patient sampled at both the SWM and WM stages (Table 1). In-house data from BM immune cell samples obtained from HD ($n=23$) and patients with SMM ($n=26$) was also integrated^{36,37}. Overall, we annotated 209,727 immune cells (T, $n=95,754$; NK, $n=24,219$; Myeloid, $n=53,172$; progenitors, $n=36,582$), excluding 28,991 normal B cells which were processed separately (Fig. 1A, Supplementary Fig. 1A). The total number of samples with at

least 100 immune cells, considered for downstream analyses, was 70 (HD: 18; IgM MGUS: 6; SWM: 19; WM: 3; SMM: 24).

Despite their early stage, patients with AWM showed significantly altered BM immune cell composition, including increased proportions of CD56^{dim} NK cells, S100B⁺ CD56^{dim} NK cells, CD56^{br} NK cells, CD16⁺ Monocytes, CD4⁺ and CD8⁺ central memory T cells (TCM), Th1 cells, Tregs, GZMB⁺ CD8⁺ TEMs, KIR⁺ CD8⁺ TEMs, and Tgd, and decreased proportions of cytokine-expressing (IL1B⁺) CD14⁺ monocytes, plasmacytoid dendritic cells (pDCs), canonical dendritic cells type 1 (cDC1), activated and interferon-stimulated T and NK cells (two-sided Wilcoxon, $q < 0.1$) (Fig. 1B). This aligns with our observations in patients with SMM (Fig. 1C), who also showed significant alterations in BM immune cell composition (two-sided Wilcoxon, $q < 0.1$) despite the absence of symptoms^{36–38}. Strikingly, patients with IgM MGUS, who had no morphological evidence of BM infiltration but had monoclonal IgM protein in their serum and/or a *MYD88* L265P mutation detected, already showed significantly higher proportion of CD56^{dim} NK cells, S100B⁺ CD56^{dim} NK cells, Th1 cells, and KIR⁺ CD8⁺ TEMs, and lower proportion of cytokine-expressing CD14⁺ monocytes, pDCs, and activated and interferon-stimulated T and NK cells compared to HD (two-sided Wilcoxon, $q < 0.1$), suggesting that certain alterations in immune cell composition may even precede morphological evidence of BM infiltration (Fig. 1D). In fact, most changes in immune cell composition could be observed as early as at the MGUS stage, suggesting that immune dysregulation is established early in the course of disease (Fig. 1E). Notably, the proportion of Tregs was significantly increased in patients with SWM compared to those with IgM MGUS (two-sided T-test, $p = 0.014$; two-sided Wilcoxon, $p = 0.092$) (Fig. 1F), in line with a prior study showing that WM tumor cells are capable of inducing Treg differentiation and expansion³⁴. This suggests that Tregs may play a role in disease progression from IgM MGUS to overt WM.

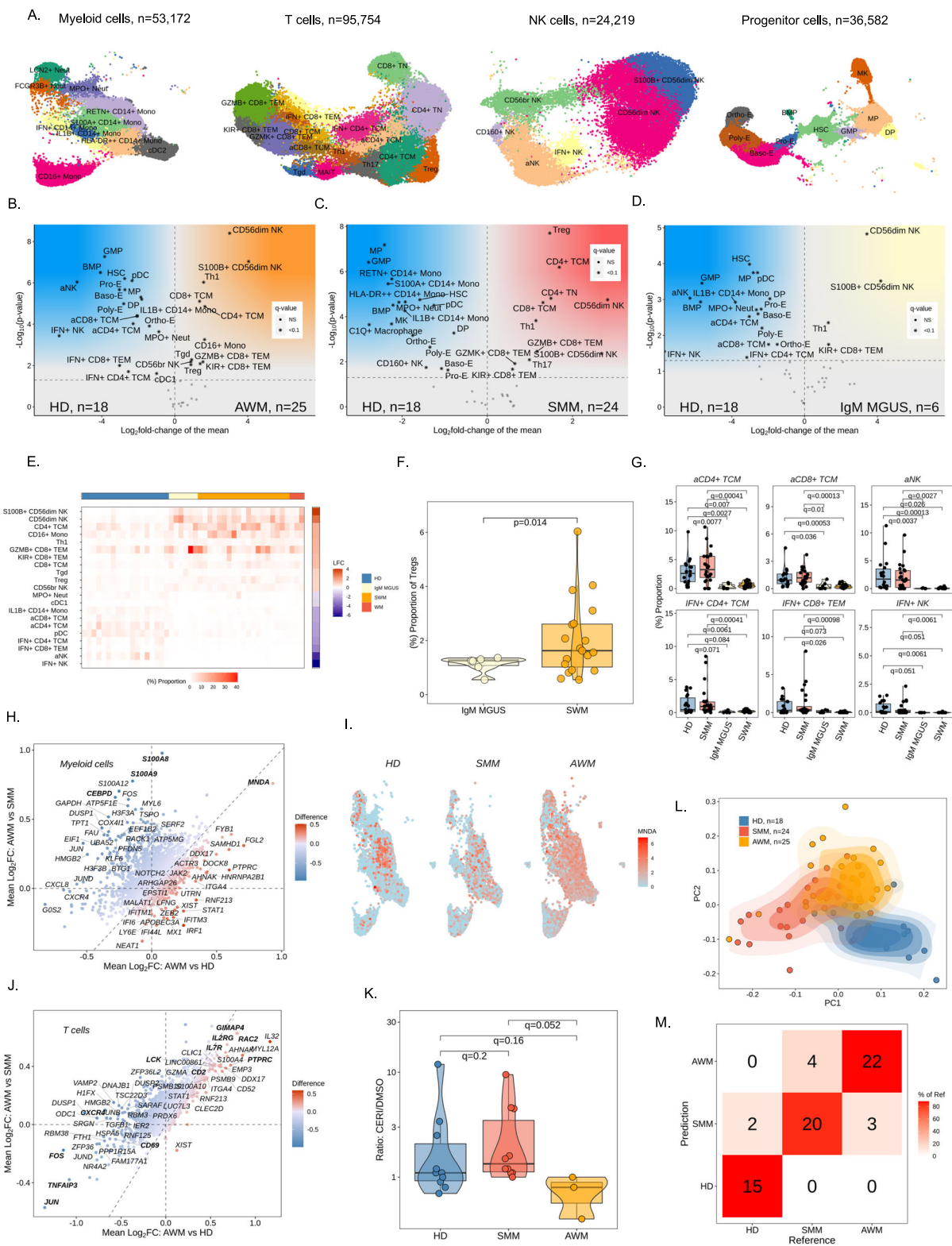
Disease-specific immune hallmarks of AWM and SMM

Next, we explored differences in the composition of the BM immune microenvironment between patients with AWM and SMM, another plasma cell premalignancy of the BM, reasoning that while some changes may reflect a general immune response against clonal expansion in the BM, others may be specific to the tumor type. In both tumor types, patients showed significantly higher proportions of S100B⁺ CD56^{dim} NK cells and CD16⁺ monocytes, however, this increase was more pronounced in patients with AWM compared to SMM (two-sided Wilcoxon, $q = 9.9 \times 10^{-4}$ and $q = 1.5 \times 10^{-2}$, respectively) (Supplementary Fig. 1B). Patients with AWM had significantly fewer activated and interferon-stimulated T and NK cells compared to patients with SMM and HD (two-sided Wilcoxon, $q < 0.1$), suggesting this may be an immune hallmark of WM (Fig. 1G, Supplementary Fig. 1B–D). Furthermore, patients with SMM had significantly fewer CD14⁺ monocytes and significantly more Tregs and naïve CD4⁺ T cells compared to both patients with AWM and HD (Supplementary Fig. 1B, Fig. 1C), suggesting these changes may be immune hallmarks of Multiple Myeloma (MM).

Disease-specific changes were also observed at the gene expression level. For example, myeloid cells from patients with SMM showed a disease-specific decrease in the expression of *CEBPD* (two-sided Wilcoxon compared to HD, $q = 8.98 \times 10^{-7}$), a transcription factor that regulates the expression of the alarmins *S100A8* and *S100A9*, which also showed disease-specific downregulation ($q = 0.012$ and $q = 0.022$, respectively) (Fig. 1H, Supplementary Fig. 1E, F)³⁹. On the other hand, myeloid cells from patients with AWM showed a disease-specific increase in the expression of *MNDA* (two-sided Wilcoxon compared to HD, $q = 5.4 \times 10^{-10}$), a key regulator of interferon response (Fig. 1H, I)^{40–44}. This change may be associated with the observed depletion of interferon-stimulated T and NK cells in patients with AWM. Furthermore, T cells from patients with AWM showed consistently higher levels of expression of genes important for receptor-mediated activation, such as *PTPRC*, *CD2*, *LCK*, *IL2RG*, *IL7R*, *GIMAP4*, and *RAC2*, along

Table 1 | Patient & sample characteristics table

PatientID	SampleID	Tumor Sample	Immune Sample	Sample Processing Method	Age	Diagnosis	Stage	BM Infiltration	Serum IgM (mg/dL)	Albumin (g/dL)	Beta-2 Microglobulin (mg/L)	MYD88	CXCR4
Pt1	S1	1	1	Ficoll	80-84	AWM	SWM	0.4	1570	4.3	2.23	L265P	WT
Pt2	S2	1	1	Ficoll	65-69	AWM	SWM	0.3	2070	4.2	2.05	L265P	WT
Pt3	S3	1	1	Ficoll	70-74	WM	WM	0.2	1170	NA	2.36	L265P	Mut
Pt4	S4	1	1	Ficoll	55-59	WM	WM	0.08	342	4.8	NA	WT	WT
Pt5	S5	1	1	Ficoll	70-74	WM	WM	0.1	944	4.5	NA	L265P	WT
Pt6	S6	0	1	Ficoll	70-74	AWM	SWM	0.1	544	4.4	1.88	L265P	WT
Pt7	S7	1	1	Ficoll	65-69	AWM	SWM	0.1	1880	4.6	1.4	L265P	NA
Pt8	S8	1	1	RBCLB	80-84	AWM	SWM	0.9	241	2.8	2.6	L265P	Mut
Pt9	S9	0	1	Ficoll	60-64	AWM	IgM MGUS	0	290	4	2.7	WT	WT
Pt10	S10	0	1	Ficoll	80-84	AWM	IgM MGUS	0	3958	2.9	3.4	WT	WT
Pt11	S11	1	1	Ficoll	80-84	AWM	SWM	0.7	1915	3.9	1.8	L265P	WT
Pt12	S12	1	1	RBCLB	65-69	AWM	SWM	0.6	2548	4.3	2.5	L265P	WT
Pt13	S13	1	1	Ficoll	70-74	AWM	SWM	0.5	1882	4	3	L265P	WT
Pt14	S14	1	1	Ficoll	50-54	AWM	SWM	0.5	3148	4.6	2.2	L265P	Mut
Pt15	S15	1	1	RBCLB	70-74	AWM	SWM	0.4	1629	4	6.7	L265P	WT
Pt16	S16	1	1	Ficoll	50-54	AWM	SWM	0.4	3574	4.2	2.1	L265P	WT
Pt17	S17	0	1	RBCLB	65-69	AWM	SWM	0.15	558	4.1	1.9	L265P	WT
Pt18	S18	1	0	RBCLB	70-74	WM	WM	0.2	1452	4.8	5.1	L265P	WT
Pt19	S19	1	1	Ficoll	65-69	AWM	SWM	0.2	755	4.3	2.8	L265P	WT
Pt20	S20	1	1	RBCLB	70-74	AWM	SWM	0.15	877	3.8	1.8	L265P	WT
Pt21	S21	1	0	RBCLB	60-64	AWM	SWM	0.8	2400	3.9	2.5	L265P	Mut
Pt22	S22	0	1	RBCLB	60-64	AWM	IgM MGUS	0	216	4.4	1.5	L265P	WT
Pt23	S23	0	1	RBCLB	65-69	AWM	IgM MGUS	0	183	4.5	2.54	WT	WT
Pt24	S24	0	1	RBCLB	70-74	AWM	SWM	0.05	1331	3.8	1.9	L265P	WT
Pt25	S25	0	1	RBCLB	50-54	AWM	IgM MGUS	0	209	4	1.5	L265P	WT
Pt26	S26	1	1	Ficoll	60-64	AWM	SWM	0.1	706	4.1	1.9	L265P	WT
Pt27	S27	1	0	RBCLB	55-59	AWM	SWM	0.1	701	4.6	1.7	L265P	WT
Pt28	S28	0	1	RBCLB	60-64	AWM	SWM	0.06	1026	4.7	2.4	L265P	WT
Pt29	S29	1	1	Ficoll	70-74	AWM	SWM	0.7	4475	4.6	1.9	L265P	Mut
Pt30	S30	1	1	Ficoll	45-49	AWM	SWM	0.7	3691	4	1.8	L265P	Mut
Pt31	S31	1	1	Ficoll	55-59	AWM	IgM MGUS	0	241	4.2	2	L265P	WT



with lower levels of common activation markers, such as *CD69*, *CXCR4*, and members of the AP-1 and NF κ B pathway (Fig. 1J, Supplementary Fig. 1G). This expression profile is consistent with the observed depletion of activated T and NK cells in patients with AWM and suggests that patient T cells may present altered activation potential compared to both HD and patients with SMM. To explore whether

T cells from patients with AWM are capable of being activated in a T Cell Receptor (TCR)-mediated manner, we stimulated peripheral blood mononuclear cells (PBMCs) from 3 patients with AWM with the CERI (CMV/EBV/RSV/Influenza A) peptide pool in vitro and measured IFN- γ levels by ELISPOT. We observed no significant difference in IFN- γ levels post-stimulation in patients with AWM compared to HD ($n = 10$;

Fig. 1 | Immune dysregulation is established early in the course of WM with disease-specific immune hallmarks. **A** Uniform manifold approximation and projection (UMAP) embeddings of myeloid cells ($n = 53,172$), T cells ($n = 95,754$), NK cells ($n = 24,219$), and progenitor cells ($n = 36,582$). **B–D** Volcano plot of proportion changes in the BM of patients with AWM (**B**, $n = 25$), SMM (**C**, $n = 24$), and IgM MGUS (**D**, $n = 6$), compared to HD ($n = 18$). Patients with at least 100 immune cells were included. P -values were computed with two-sided Wilcoxon's rank-sum test and corrected using the Benjamini-Hochberg approach. **E** Heatmap of cell type proportions in patients with IgM MGUS ($n = 6$, light yellow), SWM ($n = 19$, orange), WM ($n = 3$, red), and HD ($n = 18$, blue) (x-axis). Patients with at least 100 immune cells were included. Cell types that changed significantly between patients with AWM and HD, excluding progenitor cells, were visualized (y-axis) and the $\log_2$ fold-change was depicted in a bar (right). **F** Box plots, violin plots, and scatter plots visualizing the proportion of Tregs in patients with IgM MGUS ($n = 6$) compared to SWM ($n = 19$). Violin outline width represents density. Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times$ interquartile range (IQR). The p -value was computed with two-sided T-test. **G** Box plots, violin plots, and scatter plots comparing the proportion of activated and IFN-stimulated T and NK cells in the BM microenvironment of HD ($n = 18$), patients with SMM ($n = 24$), IgM MGUS ($n = 6$), and SWM ($n = 19$). Violin outline width represents density. Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times$ IQR. P -values were computed with two-sided Wilcoxon's rank-sum test and corrected with the Benjamini-Hochberg approach. **H** Scatter plots of genes differentially expressed in patients with AWM compared to HD (x-axis) and in patients with AWM compared to patients with SMM (y-axis) in myeloid cells. Genes

whose $\log_2$ fold-change between patients with AWM and HD is higher than that between AWM and SMM are colored in red; genes whose $\log_2$ fold-change is lower are colored in blue; genes discussed in the main text are presented in bold. **I** UMAP embedding of myeloid cells in HD ($n = 15,957$), and patients with SMM ($n = 5634$) and AWM ($n = 23,521$) colored by *MNDA* expression levels. **J** Scatter plots of genes differentially expressed in patients with AWM compared to HD (x-axis) and in AWM compared to SMM (y-axis) in T cells. Genes whose $\log_2$ fold-change between patients with AWM and HD is higher than that between AWM and SMM are colored in red; genes whose $\log_2$ fold-change is lower are colored in blue; genes discussed in the main text are presented in bold. **K** Box plots, violin plots, and scatter plots of IFN- γ expression measured by ELISPOT in PB T cells from patients with AWM ($n = 3$), SMM ($n = 10$), and HD ($n = 10$) post-stimulation with CERI peptides in vitro compared to DMSO. Each dot represents the mean of two replicates from an individual patient. Violin outline width represents density. Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times$ IQR. P -values were computed with two-sided Wilcoxon's rank-sum test and corrected with the Benjamini-Hochberg approach. **L** Principal component embedding of BM samples from patients with AWM ($n = 25$), SMM ($n = 24$), and HD ($n = 18$). Patients with at least 100 immune cells were included in this analysis. **M** Confusion matrix visualizing the accuracy of an SVM classifier trained using 5-fold cross-validation on BM samples from patients with AWM ($n = 25$), SMM ($n = 24$), and HDs ($n = 17$). Patients with at least 100 immune cells excluding progenitor cells were included in this analysis. The classifier's performance in the held-out subsets ($n = 66$) was visualized. Source data are provided as a Source Data file.

two-sided Wilcoxon, $q = 0.02$), although their levels were significantly lower than those observed in patients with SMM ($n = 10$; $p = 0.018$, $q = 0.052$) (Fig. 1K). This observation is consistent with a prior study showing that T cells from patients with WM are capable of recognizing tumor-specific antigens and reacting by producing IFN- γ in vitro³³.

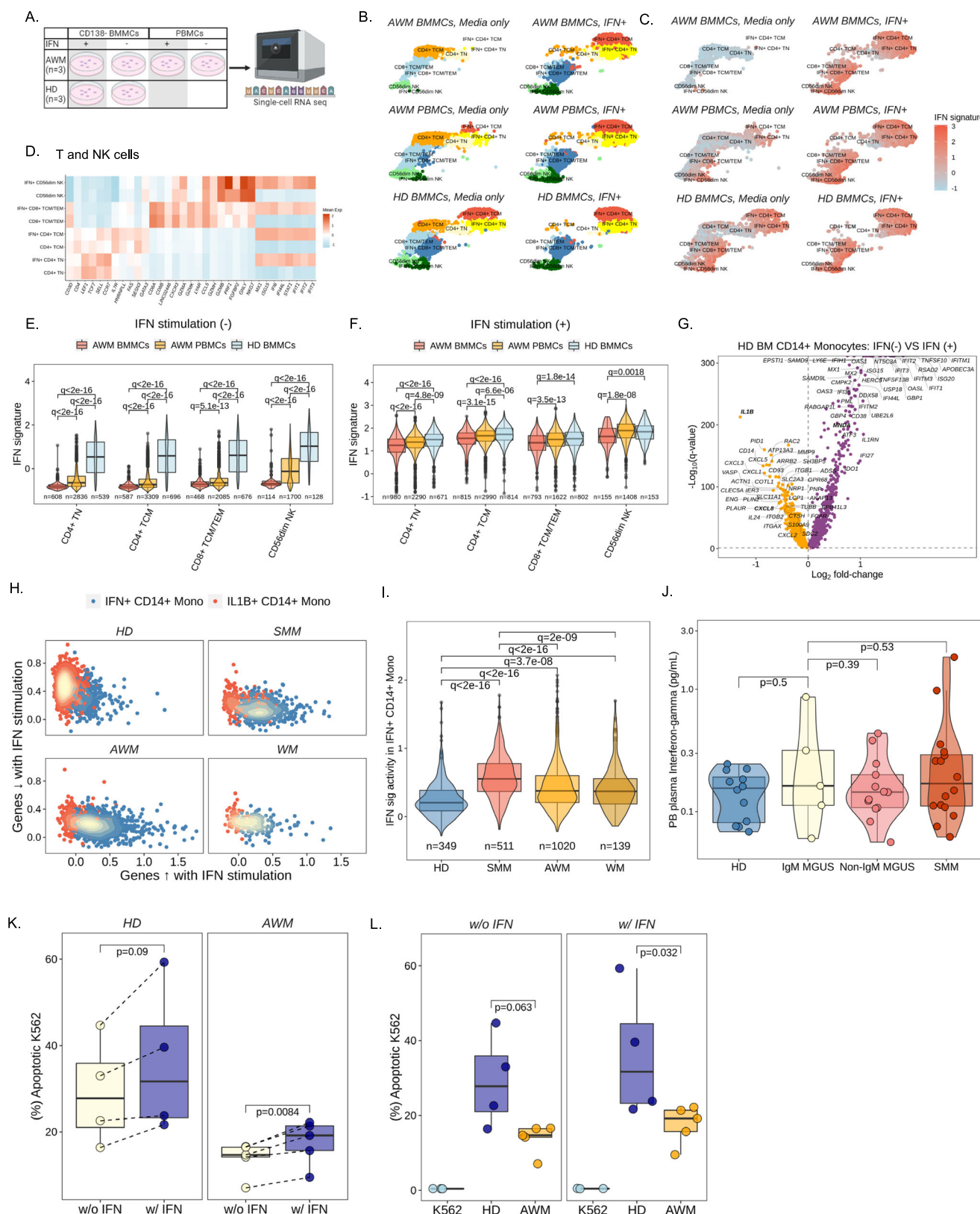
Since the two tumor types presented with different changes in immune cell composition, we reasoned that immune profiling alone may be able to diagnose and differentiate the two tumor types. Indeed, in a principal component analysis of immune cell composition, samples from patients with AWM clustered separately from both samples from patients with SMM and samples from HD (Fig. 1L). For the purposes of this analysis, we removed progenitor cells and restricted our cohort to individuals with at least 100 immune cells excluding progenitor cells (HD: $n = 17$; SMM: $n = 24$; AWM: $n = 25$). Next, we ranked features for each comparison (SMM vs HD, AWM vs HD) based on the heuristic described by Golub et al. and retained the union of the top 10 cell types across both comparisons for a total of 17 features (Supplementary Fig. 1H)⁴⁵. We randomly split the cohort into 5 subsets and performed 5-fold cross-validation, each time training an SVM classifier on the 4 training subsets and testing it on the held-out subset. Across all testing subsets ($n = 66$), the SVM classifier was able to correctly diagnose the presence of malignancy in all cases ($n = 49$) with a sensitivity of 100% (95% CI: 91–100) and a specificity of 88% (95% CI: 62–98) (Fig. 1M). Furthermore, it correctly diagnosed AWM in 22 out of 25 cases with a sensitivity of 88% (95% CI: 68–97) and a specificity of 90% (95% CI: 76–97) and correctly diagnosed SMM in 20 out of 24 cases with a sensitivity of 83% (95% CI: 62–95) and a specificity of 88% (95% CI: 74–96) (Fig. 1M). Notably, misclassification of AWM cases was not associated with the degree of BM infiltration (two-sided Wilcoxon, $p = 0.74$) (Supplementary Fig. 1I). These results suggest that it is possible to leverage immune profiling to diagnose cancer and potentially even differentially diagnose AWM from SMM in patients. Considering that 32% ($n = 8/25$) of our AWM cohort had little to no morphological evidence of BM infiltration by lymphoma ($< 10\%$ infiltration) without a discernible effect on our classifier's performance, this approach may represent a viable strategy in the early cancer detection and cancer diagnostics space.

T and NK cell from patients with AWM are hypo-responsive to interferon stimulation

Subsequently, we turned our attention to the depletion of IFN-stimulated T and NK cells in patients with AWM. To assess the capacity

of T and NK cells from patients with AWM to respond to IFN stimulation, we performed an in vitro stimulation experiment using 1000 U/mL of universal Type I IFN vs media alone on CD138⁺ BM mononuclear cells (BMMCs) from 3 patients with AWM and 3 HD and performed scRNA-seq to measure transcriptional changes (Fig. 2A). Furthermore, we included PBMCs from the same 3 patients with AWM to examine whether the potential defect was specific to the tumor site or systemic (i.e., present in both the BM and the PB). We found that the lack of IFN-stimulated cells in media-only samples from patients with AWM was systemic, although the effect was significantly more pronounced in the BM compared to the PB (two-sided Wilcoxon, $q < 0.1$) (Fig. 2B–E). Furthermore, in vitro stimulation resulted in the generation of IFN-stimulated T and NK cells in both HD and patients with AWM, however, the levels of IFN stimulation did not entirely normalize, reproducing the gradient observed in the control samples: patient BM $<$ patient PB $<$ HD BM (two-sided Wilcoxon, $q < 0.1$) (Fig. 2F). This suggests that T and NK cells from patients with AWM may be hypo-responsive to IFN stimulation. Notably, the dose used for in vitro stimulation (1000 U/mL) was selected based on a prior study showing a measurable impact at the RNA level for both IFN- α and IFN- β -stimulated genes at this dose⁴⁶. Nevertheless, this dose is ~ 30 – $300\times$ times higher than the concentration typically encountered inside the human body (~ 3.5 ng/mL, compared to 10–100 pg/mL in healthy blood) which may explain why the observed defect was largely overcome^{47–49}.

To confirm that the observed defect was not due to the lack of IFN in the BM microenvironment of patients with AWM, we compared the gene expression profile of HD BM CD14⁺ Monocytes between the IFN-stimulated and control samples. As expected, we observed significant upregulation of key IFN response genes, including *MNDA*, which we showed before is specifically upregulated in monocytes from patients with AWM compared to both HD and patients with SMM (Fig. 2G). Moreover, we observed downregulation of pro-inflammatory monocyte markers, such as *IL1B* and *CXCL8* (encoding IL-8), corresponding to a subpopulation which we showed before is depleted in both patients with AWM and patients with SMM compared to HD (Fig. 2G). Next, we defined two signatures, comprising the top 40 genes that were significantly upregulated with the highest $\log_2$ fold-change in HD monocytes post-IFN stimulation and the top 40 genes that were significantly downregulated with the highest $\log_2$ fold-change (two-sided Wilcoxon, $q < 0.05$). We then compared the levels of these signatures in BM CD14⁺ monocytes from HD, patients with AWM/WM, and



patients with SMM. Consistent with prior knowledge that patients with SMM/MM present increased interferon signaling, we observed a pronounced shift in monocyte density towards the IFN-stimulated profile in BM monocytes from patients with SMM (Fig. 2H).^{37,50,51} Notably, however, we observed a similar shift in patients with AWM and WM whose monocytes showed significantly higher activity of the IFN signature compared to HD (two-sided Wilcoxon, $q < 0.1$), although the

activity was lower in patients with AWM compared to SMM (Fig. 2H, I). This suggests that IFN levels in the patient BM are not abnormally low. Next, we measured the levels of IFN- γ in PB plasma from HD ($n = 13$), patients with IgM MGUS ($n = 5$), non-IgM MGUS ($n = 15$), and SMM ($n = 20$), using Olink's proximity-extension assay (Fig. 2J). We could not find a significant change in IFN- γ levels between patients with IgM MGUS and either HD or patients with non-IgM MGUS/SMM (two-sided

Fig. 2 | T and NK cells from patients with AWM are hypo-responsive to interferon stimulation. **A** Experimental design. CD138⁺ BMMCs from patients with AWM ($n = 3$) and HD ($n = 3$) were cultured with or without 1000 U/mL of universal type I IFN and subjected to scRNA-seq to measure transcriptional changes post-stimulation. PBMCs from the same 3 patients with AWM were also examined (Created in BioRender. Ghobrial, I. (2024) BioRender.com/i72j359). **B**, **C** UMAP embeddings of T and NK cells ($n = 27,239$) labeled by subtype (**B**) or colored by the activity of an IFN stimulation signature (**C**). **D** Heatmap of gene expression markers (scaled mean expression) used for cell type annotation. **E**, **F** Box plots and violin plots of the level of IFN stimulation signature (y-axis) in subtypes of T and NK cells (x-axis) from patients with AWM (BMMCs: red; PBMCs: orange) and HD (BMMCs: light blue) following in vitro culture without (**E**) or with type I IFN stimulation (**F**). Each dot represents a cell. Violin outline width represents density. Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$. *P*-values were computed with two-sided Wilcoxon's rank-sum test and corrected with the Benjamini-Hochberg approach. **G** Volcano plot of gene expression changes in CD14⁺ monocytes from the BM of HD ($n = 3$) post-stimulation with type I IFN. Each dot represents a gene (upregulated: purple; downregulated: orange). **H** Scatter plot and density plot of IFN stimulation gene signatures in IFN⁺ CD14⁺ monocytes (blue) and IL1B⁺ CD14⁺ monocytes (red)

from patients with AWM, WM, SMM, and HD. Each dot is a cell. The x- and y-axes correspond to the signature scores defined by genes increased and decreased post-IFN stimulation, respectively. **I** Box plots and violin plots of the level of IFN stimulation signature in IFN⁺ CD14⁺ monocytes from patients with SMM (red), AWM (orange), WM (dark orange), and HD (blue). Each dot is a cell. Violin outline width represents density. Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$. *P*-values were computed with two-sided Wilcoxon's rank-sum test and corrected with the Benjamini-Hochberg approach. **J** Box plots, violin plots, and scatter plots of IFN- γ concentration (pg/mL) in PB plasma samples from patients with IgM MGUS ($n = 5$), non-IgM MGUS ($n = 15$), and SMM ($n = 20$), compared to HD ($n = 13$). Violin outline width represents density. Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$. *P*-values were computed with two-sided Wilcoxon's rank-sum test and corrected with the Benjamini-Hochberg approach. **K**, **L** Boxplots and scatter plots of the proportion of apoptotic K562 cells following co-culture with NK cells from patients with AWM ($n = 5$) and HD ($n = 4$) with or without the administration of IFN-I (1000 U/mL). Dashed lines connect samples from the same individual. Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$. *P*-values were computed with two-sided paired T-test (**K**) or Wilcoxon's rank-sum test (**L**). Source data are provided as a Source Data file.

Wilcoxon, IgM MGUS vs HD, $p = 0.5$; IgM MGUS vs non-IgM MGUS, $p = 0.39$; IgM MGUS vs SMM, $p = 0.53$; $q = 0.53$ for all comparisons). Collectively, these results suggest that despite the presence of IFN in the circulation and the BM microenvironment, T and NK cells from patients with AWM are hypo-responsive to IFN stimulation, a defect that may be overcome by exogenous IFN administration.

Two small previous studies recorded some degree of clinical response to therapeutic IFN administration in certain patients with WM, which suggests that this defect may translate into a therapeutic vulnerability in some patients^{52,53}. To pursue this hypothesis, we compared the cytotoxicity potential of PB-derived NK cells between patients with AWM ($n = 5$) and HD ($n = 4$), with or without the administration of 1000 U/mL of IFN-I. NK cell cytotoxicity was assessed by measuring the proportion of apoptotic K562 cells via flow cytometry following co-culture of NK cells with K562 cells (Supplementary Fig. 2A). We observed a significant increase in NK cell cytotoxicity in patients with AWM (Paired t-test, $p = 0.0084$) with the administration of IFN-I, however NK cells from patients with AWM showed significantly lower cytotoxicity potential compared to those from HD post-IFN stimulation (Wilcoxon, $p = 0.032$) (Fig. 2K, L). Notably, a similar trend could be observed at baseline ($p = 0.063$), suggesting that NK cells from patients with AWM are indeed dysfunctional. These results are concordant with the results of our single-cell RNA-seq experiment and demonstrate that NK cell functionality is impaired in patients with AWM compared to healthy donors, and that IFN administration can improve, although not entirely correct, NK cell functionality in patients, potentially supporting a role for IFN administration in patients with WM.

Single-cell RNA and BCR sequencing resolves concurrent lymphoid clones

Next, to dissect tumor-intrinsic genomic and transcriptomic risk factors of progression, we performed scRNA-seq and scBCR-seq on CD138⁺ and/or CD19⁺ tumor cells from 18 patients with IgM MGUS/SWM (IgM MGUS: $n = 1$; SWM: $n = 17$) and 4 patients with WM, including 1 patient sampled at both stages (Table 1). Overall, we identified 48,875 WM tumor cells in 21 out of 22 samples profiled (median: 1089, range: 106–8694) (Fig. 3A, B, Supplementary Fig. 3A, B). One patient (Pt5) had a second sample drawn five years later at the WM stage, following 5 cycles of treatment [AWM timepoint (T1): S12, WM timepoint (T2): S5]. All but one patient (Pt4), who had low BM infiltration, were positive for *MYD88* L265P. Tumor cells from the patient of unknown *MYD88* status clustered with the rest of the tumors, suggesting that they are likely positive for the mutation, which can be missed in the presence of low BM infiltration (Fig. 3A). One patient (Pt19), who was clinically thought

to have WM with Chronic Lymphocytic Leukemia (CLL)-like immunophenotype due to the presence of *MYD88* L265P in the context of CD5⁺ B cell lymphoproliferative disease in the BM, had no WM tumor cells detected; instead, an unclassified lymphoma of undetermined type (UCL) and two Monoclonal B Cell Lymphocytosis (MBL)/CLL clones were detected, whose immunoglobulin isotypes (IgM Kappa) did not match the (small) IgM Lambda M-spike observed on immunofixation. This suggests that this patient has multiple lymphoid pre-malignancies and that the small IgM-producing and (likely) *MYD88*-mutant clone may be unrelated to the CD5⁺ B cells detected on the BM biopsy. Three other patients (Pt1, Pt8, Pt14) showed concurrent (clinically undetected) MBL/CLL clones, bringing the total number of cases with concurrent MBL/CLL to 4 and the overall frequency of concurrent MBL/CLL to 19% ($n = 4/21$, 95% CI: 6–43) (Fig. 3A, B). Notably, two of these four patients (Pt8, Pt19) had more than two lymphoid clones each, suggesting that these patients may be predisposed to developing lymphoid clones. Two patients (Pt3, Pt8) showed biclonal WM, and one patient (Pt13) showed triclonal WM, for an overall frequency of multiclonal WM of 14% ($n = 3/21$, 95% CI: 4–37) (Supplementary Fig. 3A). Patient Pt8, who presented clinically with IgM and IgG M-spikes on immunofixation and who was thought to have IgG LPL with an IgM⁺ component, had two distinct WM tumors, one of which was IgM⁺ while the other was IgG⁺; this observation could be valuable for properly monitoring clonal dynamics and response to therapy over time⁵⁴. The other two patients with multiclonal disease had a single M-spike on immunofixation. While no systematic analysis of multiclonal disease has been published for patients with WM, biclonal gammopathies are rare by immunofixation (~1.6% of all gammopathies) and frequently represent clonally related populations rather than true multiclonal disease as is the case with our patients^{55–57}. Collectively, these findings highlight the discriminatory power of scRNA/BCR-seq in clinical diagnostics.

Increased abundance of marginal zone-like, IgE⁺, atypical, and class-switched memory B cells in patients with AWM

A recent study utilizing mass cytometry has shown that non-clonal B cells from patients with IgM MGUS and WM are enriched for CXCR5-negative extrafollicular B cells, however the normal B cell compartment of patients with AWM has not been studied at a large scale before³³. The use of scBCR-seq coupled with the use of samples from early-stage patients who tend to have a lower degree of BM infiltration allowed us to capture and profile 28,991 normal B cells (Supplementary Fig. 3C, D). Specifically, we resolved a CD20⁺ CD38⁺ transitional B cell population (TBC); IL4R⁺ ABCB1⁺ naïve B cells (NBC); CD27⁺ IgM⁺ non-class-switched memory B cells (NCSMB); CD1C⁺ PLD4⁺ marginal

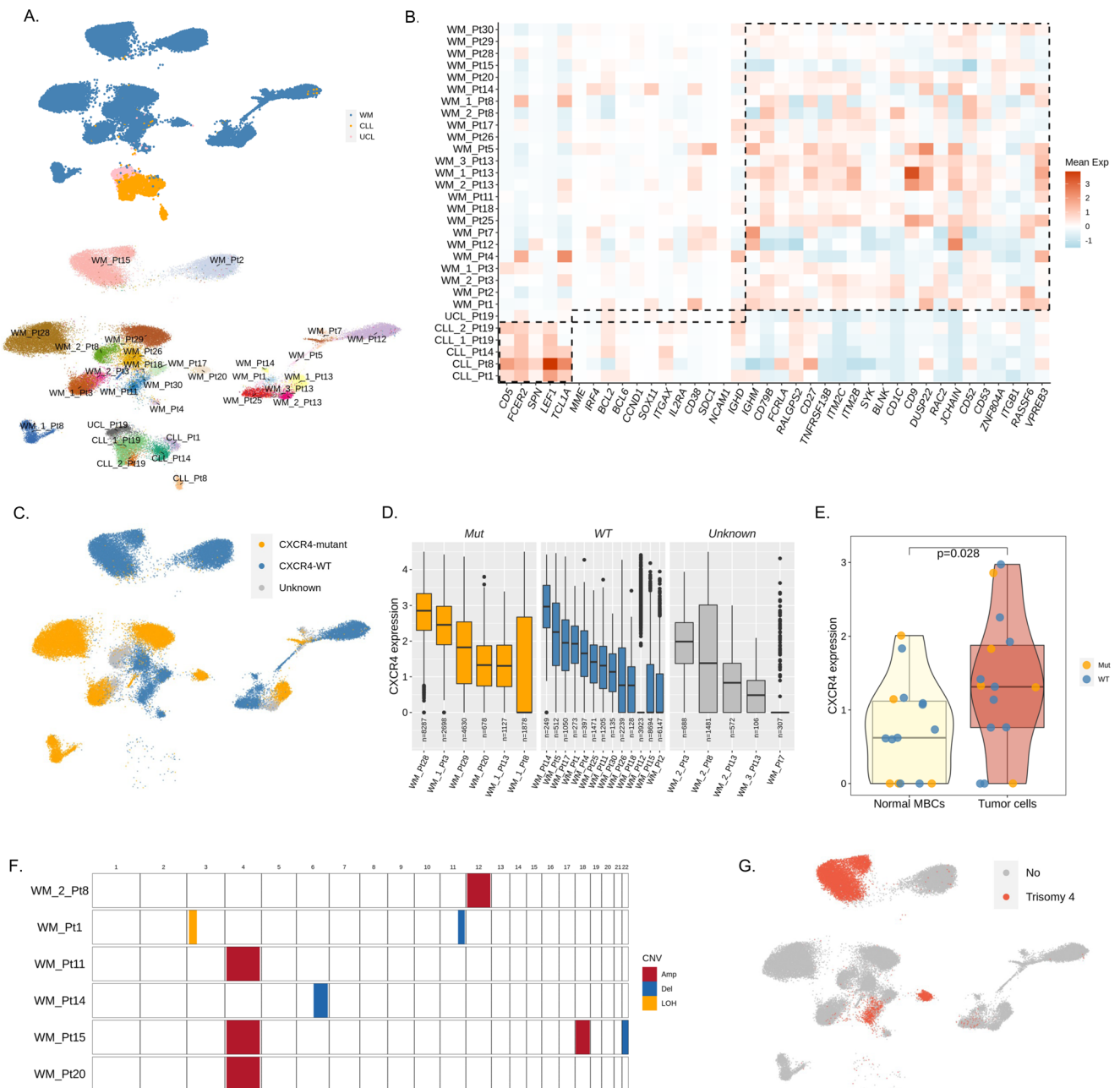


Fig. 3 | Single-cell RNA and BCR sequencing resolves concurrent lymphoid clones and unveils inter-tumor heterogeneity. **A** UMAP embeddings of tumor cells ($n = 55,488$) labeled by tumor type (top) or tumor instance (bottom). Sample notation: Tumor Type_Tumor Number_Patient Number (for example: WM_1Pt8 corresponds to the 1st of 2 WM tumors detected in patient Pt8). Tumor number is only noted when multiple tumors of that type exist in that patient. **B** Heatmap of tumor type-specific gene expression markers (scaled mean expression) used to determine tumor type. **C** UMAP embedding of WM tumor cells ($n = 48,875$) colored by *CXCR4* mutation status. **D** Box plots visualizing the level of *CXCR4* expression (y-axis) in each WM tumor clone (x-axis). Tumors were grouped into three categories: *CXCR4*-mutant ($n = 6$), *CXCR4*-WT ($n = 13$), and unknown *CXCR4* mutation status ($n = 5$). Tumors were sorted by median *CXCR4* expression in decreasing order. Box:

1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$. **E** Box plots, violin plots, and scatter plots of median *CXCR4* expression (y-axis) in WM tumor cells compared to normal memory B cells (MBCs) from the same patients ($n = 15$) (x-axis). Patients without tumor or normal B cells ($n = 5$) were removed from this analysis. In patients with multiclonal WM, the primary clone was used to calculate median *CXCR4* expression. Patients were colored by *CXCR4* mutation status. Violin outline width represents density. Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$. The p -value was computed with a paired two-sided Wilcoxon's rank-sum test. **F** Heatmap of CNVs detected in tumor samples. **G** UMAP embedding of WM tumor cells ($n = 48,875$) colored by Trisomy 4 status. Source data are provided as a Source Data file.

zone-like B cells (MZBC); activated B cells expressing genes of the AP-1 pathway (aBC); T-bet⁺ CD11c⁺ atypical B cells (ABC); a TCF4⁺ class-switched memory B cell population which co-expressed *RASSF6* and *JCHAIN* and showed higher levels of IgA expression (TCF4⁺ MBC); IgE⁺ memory B cells, co-expressing *IGHG1* and *IGHG4* (IgE⁺ MBC); and two class-switched memory B cell populations, one expressing higher

levels of *TCF7* (TCF7⁺ MBC) and another one expressing higher levels of *S100A10*, *CD99*, and *ITGB1* (S100A10⁺ MBC).

By comparing B cell subtype proportions (out of all B cells) between patients with AWM ($n = 13$) and HD ($n = 10$) with at least 50 B cells captured in total, we observed a significant switch toward memory B cells, with a decreased abundance of naïve B cells (two-sided

Wilcoxon, $q = 1.4 \times 10^{-4}$) (Supplementary Fig. 3E). Furthermore, we observed a significant decrease in activated B cells ($q = 1.4 \times 10^{-4}$), which is in line with our data in the T and NK cell compartments and suggests a more general defect in the activation of lymphocytes may be present in patients with WM. We observed a significant increase in marginal zone-like B cells ($q = 2.1 \times 10^{-3}$), an IgM⁺ subpopulation that resembles WM tumor cells phenotypically and may represent the cell type of origin for WM. We also observed a significant increase in IgE⁺ memory B cells ($q = 3.5 \times 10^{-5}$), which could be associated with the increased presence of mast cells in the BM of patients with WM; higher numbers of mast cells are associated with suboptimal outcomes and support the growth of WM cells^{58,59}. Lastly, we observed a significant ($q < 0.1$) increase in the proportion of class-switched memory B cells (including the TCF4⁺, TCF7⁺, and S100A10⁺ subpopulations), as well as atypical B cells ($q = 4.8 \times 10^{-4}$), and a significant decrease in the proportion of non-class-switched memory B cells ($q = 2.9 \times 10^{-3}$). These results suggest that the normal B cell compartment in patients with AWM presents extensive changes, which could be contributing to disease progression or potentially even oncogenesis.

Single-cell RNA sequencing unveils unappreciated inter-tumor heterogeneity within MYD88-mutant WM

While each tumor clustered separately from the rest, WM tumor cells appeared to form distinct groups, suggestive of biological heterogeneity within MYD88 L265P-mutant WM which is currently considered a single entity, only subset into CXCR4-mutant vs wild-type tumors³. In our cohort of 21 patients with WM tumor cells detected, 6 patients were positive for CXCR4 mutations, 14 patients were wild-type, and 1 patient (Pt7) had no leftover cells for testing⁶⁰. Clinically undetected secondary WM tumors in patients Pt3, Pt8, and Pt13, were considered of unknown status. While CXCR4 mutation status appeared to explain part of the heterogeneity observed in the cohort, CXCR4-mutant tumors did not all cluster together or separately from CXCR4-WT ones (Fig. 3C). The impact of CXCR4 mutations on CXCR4 expression has not been studied extensively, and is challenging to resolve due to native CXCR4 expression in normal B cells which often contaminate bulk profiling analyses, as well as possible false negative results in CXCR4 mutation testing related to the subclonal frequency of CXCR4 mutations and the degree of tumor infiltration in the BM^{61,62}. A prior study reported no increase in CXCR4 expression levels in CXCR4-mutant patients using bulk RNA-seq⁶¹. By separating malignant and normal B cells based on scBCR-seq, we observed that while CXCR4 expression was high in CXCR4-mutant patients, high levels of CXCR4 expression could also be seen in CXCR4-WT tumors as well (Fig. 3D). This was also true when attempting to control for potential false negative results by considering CXCR4-WT patients separately if they had high BM infiltration (30% or higher) (Supplementary Fig. 3F). Notably, CXCR4 levels were significantly higher in WM tumor cells compared to normal memory B cells (MBCs) from the same patients (paired two-sided Wilcoxon, $p = 0.028$) (Fig. 3E). These results suggest that CXCR4 expression can be upregulated in WM tumors independent of CXCR4 mutations.

To explore whether the observed heterogeneity could be explained by underlying copy number variants (CNVs), we used Numbat to infer CNVs in tumor cells⁶³. Overall, CNVs were detected in 6 tumors (Pt14: Del6q; Pt11: Trisomy 4; Pt15: Trisomy 4, trisomy 18, Del22q; Pt1: Del11q, 3p LOH; Pt20: Trisomy 4; secondary WM tumor of Pt8: Trisomy 12), all of which were from patients with AWM (Fig. 3F). The frequency of any CNV in patients with AWM and tumor cells detected was 35% ($n = 6/17$, 95% CI: 15–61), which is higher than previously reported (i.e., 18–20%)^{25,28}. Most CNVs detected were unique to their tumors except trisomy 4 which occurred in three different tumors ($n = 3/17$, 18%, 95% CI: 4.7–44). This frequency was significantly higher than previously reported for patients with AWM (3.8%) (two-sided binomial, $p = 0.025$), however, trisomy 4 did not appear to

explain the observed patterns of variation (Fig. 3G)²⁸. Collectively, these results suggest that additional variables, other than CXCR4 mutation and CNV status, may be shaping the transcriptional landscape of WM.

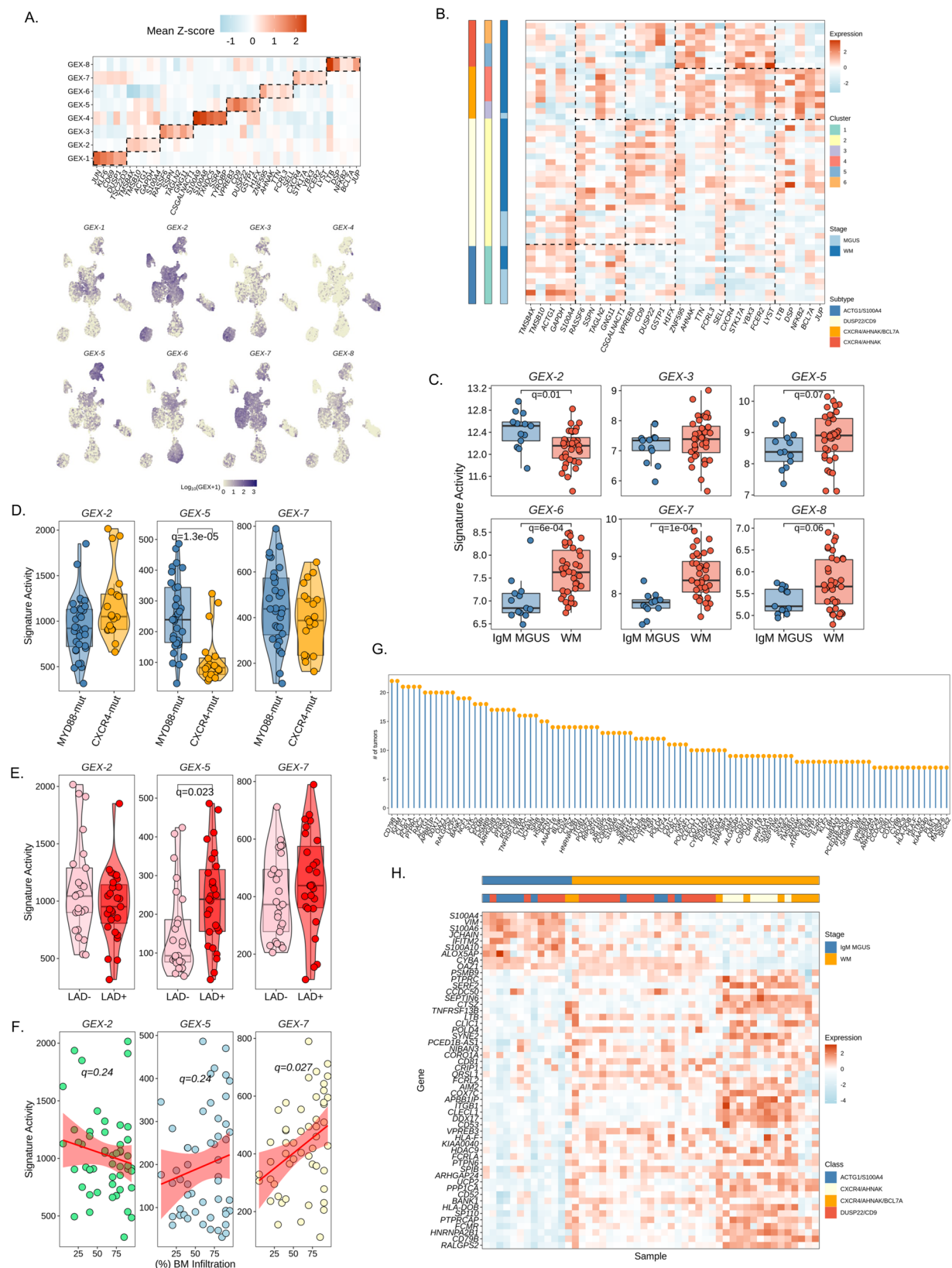
Expression-based classification of MYD88-mutant WM tumors reveals subtypes of disease and identifies expression signatures associated with progression

To investigate the observed heterogeneity, we performed Bayesian non-negative matrix factorization on a matrix of WM tumor cells using the top 2000 highly variable genes. Overall, eight gene expression signatures (GEX) were extracted, which showed variable activity across tumors (Fig. 4A). Signatures GEX-1 and GEX-4 were removed from further consideration, as they corresponded to either a generic immune cell activation module (GEX-1) or monocytic contamination (GEX-4). For the remaining six signatures, key marker genes included CXCR4; the tumor suppressor RASSF6; AHNK, a gene involved in cytokinesis; the transcription factor ZNF595; thymosins beta 4 (TMSB4X) and beta 10 (TMSB10); DUSP22, which is located next to IRF4 on chromosome 6 and is involved in structural variants in lymphoma; CD9, a gene associated with B cell apoptosis, dependence on follicular dendritic cells, and disease progression in lymphoma; and BCL7A, a gene frequently mutated in lymphoma and myeloma^{64–72}.

To assess how well these markers could delineate subtypes of WM patients, we performed hierarchical clustering on the scaled feature list of the top 5 markers from each signature ($n = 30$) in an external GEP dataset containing 13 patients with IgM MGUS and 36 patients with WM (GSE171739) (Fig. 4B)²⁷. We observed relatively consistent covariation of expression for our key markers of signatures GEX-2, GEX-5, GEX-6, GEX-7, and GEX-8, delineating four subsets of tumors: a CXCR4-high subset, which aligns with the AHNK-high subset and can be further broken down to a BCL7A-high subset; a DUSP22/CD9-high subset, which only partly overlaps with the CXCR4/AHNK-high subset; and a ACTG1/S100A4-high subset, which co-expressed thymosin beta 4 (TMSB4X) and thymosin beta 10 (TMSB10) (Fig. 4B). Interestingly, the expression of FCER2, which encodes CD23, appeared to co-segregate with the CXCR4/AHNK-high subset, suggesting that CD23 and CD9 may be useful surface markers for WM subclassification. CD23 was previously shown to be expressed in a subset of patients with WM who showed higher levels of serum IgM, however an association with CXCR4 mutations has not been described before⁷³.

Patients with IgM MGUS mostly co-segregated with the ACTG1/S100A4-high and DUSP22/CD9-high subsets (Fig. 4B). Comparison of signature activity between patients with IgM MGUS and WM revealed significantly lower activity of GEX-2 (two-sided Wilcoxon, $q = 0.01$) (corresponding to the ACTG1/S100A4-high subset) and significantly higher activity of GEX-6/GEX-7/GEX-8 ($q = 6 \times 10^{-4}$, $q = 1 \times 10^{-4}$, and $q = 0.06$, respectively) (corresponding to the CXCR4/AHNK-high subset) and GEX-5 (corresponding to the DUSP22/CD9-high subset) ($q = 0.07$) in patients with overt disease (Fig. 4C). A prior study reported a higher risk of progression in patients with CXCR4 mutations²⁰. Notably, however, signature GEX-7 was active in tumors with and without reported CXCR4 mutations, thus it may capture risk of progression in both CXCR4-mutant as well as WT tumors (Supplementary Fig. 4A). In patient Pt5, who was CXCR4-WT and who was sampled at the AWM stage (T1: S12) and then again at the WM stage five years later and following 5 cycles of treatment (T2: S5), the proportion of cells assigned to GEX-7 increased significantly from timepoint 1 to timepoint 2 (two-sided Fisher's exact, $p = 5.1 \times 10^{-8}$) (Supplementary Fig. 4B).

To further validate these signatures and assess their clinical relevance, we analyzed an external bulk RNA-seq cohort of 52 patients with MYD88-mutant WM and available genotypic information for CXCR4 mutations, as well as clinical data on the degree of BM infiltration, serum IgM levels, the presence of lymphadenopathy (LAD) and



splenomegaly⁶¹. We scored patients for our three major signatures (GEX-2, GEX-5, and GEX-7) and compared signature activity with clinical and genomic variables. In line with our prior results, we observed no significant association between the presence of a *CXCR4* mutation ($n=20$) and the activity of signature GEX-7 (two-sided Wilcoxon, $q=0.23$), which is marked by *CXCR4* (Fig. 4D). However, *CXCR4*-mutant

patients showed significantly lower activity of signature GEX-5 ($q=1.3e-05$), suggesting that patients with high activity of GEX-5 may be more likely to be *CXCR4*-WT and therefore more likely to respond better to therapy with Ibrutinib⁷⁴⁻⁷⁶. Indeed, *DUSP22* was previously shown to be significantly downregulated in *CXCR4*-mutant patients along with other dual specificity phosphatases⁶¹. Interestingly, the

Fig. 4 | Expression-based classification of *MYD88*-mutant WM tumors reveals subtypes of disease and identifies expression signatures associated with progression. **A** Heatmap of marker gene expression (mean Z-score) for gene expression (GEX) signatures active in WM tumor cells (top) and UMAP embedding of WM tumor cells ($n = 48,875$) colored by the $\log_{10}$ -scaled activity of each signature (bottom). **B** Heatmap of scaled expression for 30 GEX signature marker genes in an external GEP dataset of patients with IgM MGUS ($n = 13$) and WM ($n = 36$) (GSE171739). Vertical dashed lines delineate distinct signature markers; horizontal dashed lines delineate distinct subtypes of disease, as detected via hierarchical clustering. On the left, bars visualize the patients' disease stage, clusters, and subtypes. **C** Box plots and scatter plots visualizing the activity of each signature between patients with IgM MGUS ($n = 13$) and WM ($n = 36$) in the GEP dataset. Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$. P -values were computed with two-sided Wilcoxon's rank-sum test and corrected with Benjamini-Hochberg. **D** Box plots, violin plots, and scatter plots comparing signature activity for signatures GEX-2, GEX-5, and GEX-7 between patients with ($n = 20$) and without *CXCR4* mutations ($n = 32$) in an external bulk RNA-seq dataset. Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$. Violin outline width represents density. P -values were computed with two-sided Wilcoxon's rank-sum test and corrected with

Benjamini-Hochberg. **E** Box plots, violin plots, and scatter plots comparing signature activity for signatures GEX-2, GEX-5, and GEX-7 between patients with ($n = 27$) and without LAD ($n = 25$) in an external bulk RNA-seq dataset. Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$. Violin outline width represents density. P -values were computed with two-sided Wilcoxon's rank-sum test and corrected with Benjamini-Hochberg. **F** Scatter plot of signature activity for signatures GEX-2, GEX-5, and GEX-7 (y-axis) and BM infiltration percentage (x-axis) in an external bulk RNA-seq dataset. P -values were computed using a two-sided Pearson's correlation test and corrected with Benjamini-Hochberg. A line was fit using the "lm" method and the confidence interval is represented in light red. **G** Bar plot of genes ($n = 101$, x-axis) that are upregulated in more than 30% of WM tumors compared to memory B cells. The y axis corresponds to the number of tumors showing significant upregulation (two-sided Wilcoxon, $q < 0.05$ & $\log_2$ fold-change > 0.5) of the given gene. **H** Heatmap of the subset of Fig. 4G genes (y-axis) that were significantly up- or downregulated (two-sided Wilcoxon, $q < 0.05$) between patients with IgM MGUS ($n = 13$) and WM ($n = 36$) (x-axis) in the GEP dataset. Top bars visualize the patient's subtype and disease stage. Genes were sorted based on the difference of mean expression in patients with IgM MGUS and WM. Source data are provided as a Source Data file.

presence of LAD ($n = 27$) was associated with significantly higher GEX-5 activity (two-sided Wilcoxon, $q = 0.023$) (Fig. 4E). This is in line with signature GEX-5 being marked by *CD9*, a gene shown to denote the dependency of lymphoma cells on follicular dendritic cells, making it more likely that these cells will remain in the lymph node^{68,69}. This association may also explain why patients with WM showed higher activity of GEX-5 compared to patients with IgM MGUS, as the presence of LAD is a criterion for the diagnosis of overt disease^{7,8,77}. In contrast, no signature was associated with the presence of splenomegaly or serum level of IgM. Lastly, we observed a significant positive correlation between the activity of our high-risk signature, GEX-7, and the degree of BM infiltration (Pearson's correlation test, $q = 0.027$), suggesting that this signature may be prognostically relevant even in the setting of overt WM (Fig. 4F).

Recurrently dysregulated genes in patients with AWM and association with progression

To identify genes that are frequently dysregulated in WM tumors, we performed differential expression analysis on a per-patient level, each time comparing a patient's tumor cells to normal memory B cells. We searched for genes that were significantly (two-sided Wilcoxon, $q < 0.05$) upregulated ($\log_2$ fold-change > 0.5) in more than 30% of tumors (Fig. 4G). As expected, *IGHM* (encoding the constant region of IgM) was the most frequently upregulated gene, compared to memory B cells which contain a mixture of IgM⁺ and class-switched B cells. Similarly, *CD79B*, which is frequently mutated in patients with WM and regulates BCR signaling, was significantly upregulated in most patients, as were proximal BCR components, *SYK* and *BLNK*^{3,61}. Genes that are typically expressed in marginal zone B cells, such as *ITM2C*, *ITM2B*, *JCHAIN*, and *CD1C*, were frequently upregulated, as were genes that we previously discovered in our signature analysis, such as *DUSP22*, *RASSF6*, and *VPREB3*. In line with these results, *JCHAIN* and *ITM2B*, as well as *RASSF6* were previously shown to be upregulated in tumor cells from patients with overt WM⁷⁸. Most tumors showed significantly higher expression of *CD52*, the target of alemtuzumab, which can lead to therapeutic responses in patients with WM, and *CD53*, a surface tetraspanin that interacts with *CXCR4* to facilitate downstream signaling^{79–81}. Notably, one of the top hits was *RAC2*, a gene that interacts with phospholipase C Gamma 2 and can bypass Bruton's Tyrosine Kinase (BTK) in downstream BCR signaling, driving resistance to the BTK inhibitor ibrutinib in patients with Diffuse Large B Cell Lymphoma (DLBCL) and Mantle Cell Lymphoma (MCL)^{82,83}. It is thus possible that the high levels of *RAC2* expression in most WM tumors may be related to the development of resistance to ibrutinib in patients with WM^{74–76,84,85}. Approximately half of these genes ($n = 53$ /

101) were significantly (two-sided Wilcoxon, $q < 0.05$) up/down-regulated between patients with IgM MGUS ($n = 13$) and overt WM ($n = 36$) in the external GEP dataset, suggesting that they may be relevant for prognostication (Fig. 4H)²⁷. Strikingly, genes upregulated in patients with WM appeared to segregate patients into two broad classes, which largely overlapped with our *CXCR4/AHNAK*-high and *DUSP22/CD9*-high subsets, while genes downregulated in patients with WM aligned with our *ACTG1/S100A4*-high subset, which was enriched for patients with IgM MGUS. Genes like *ITGB1* (encoding CD29), *CLECL1*, and *SYNE2* appeared to be upregulated primarily in the *CXCR4/AHNAK*-high subset of WM cases, which suggests that further analyses of these disease subgroups may yield more markers and more insight into the underlying biology of progression.

Discussion

Understanding the impact of tumor-intrinsic transcriptomic alterations and changes in the BM immune microenvironment on disease progression from IgM MGUS to overt WM may have important implications for diagnosis, prognostication, and treatment decisions. In this study, we comprehensively characterized the tumor transcriptome and immune microenvironment of 30 patients with IgM MGUS/SWM/WM, along with 26 patients with SMM and 23 HD using single-cell RNA sequencing. Due to the relatively large number of patients on the WM disease spectrum analyzed, we were able to identify immune and gene expression biomarkers with potential implications for prognosis and treatment decisions.

Prior smaller studies have shown that the BM immune microenvironment of patients with IgM MGUS and WM may show increased proportions of CD8⁺ T cells and NK cells^{33,86,87}. We demonstrated that the BM immune cell composition of patients with AWM is indeed significantly dysregulated, suggesting that patients may have compromised immune function prior to displaying symptoms, similar to our findings in other precursor plasma cell dyscrasias^{36–38}. Notably, in patients with AWM, we observed significant changes in immune cell composition even in the absence of morphological evidence of lymphoma infiltration in the BM, which suggests that immune dysregulation may be established even earlier than previously thought. Specifically, we showed that CD16⁺ Monocytes, CD56_{dim} NK cells, S100B⁺ CD56_{dim} NK cells, CD8⁺ and CD4⁺ TCMs, GZMB⁺ CD8⁺ TEMs, KIR⁺ CD8⁺ TEMs, Tgd, Th1 cells, and Tregs are significantly more abundant in patients with AWM, who in turn exhibit depletion of pDCs, cytokine-expressing CD14⁺ Monocytes, and interferon-stimulated and activated lymphocytes. A prior study reported decreased expression of interferon-stimulation and activation markers in T cells from patients with WM, further corroborating this observation³³. Notably,

we observed variability in the ways in which the BM immune micro-environment changed across patients, suggesting that some of these changes may be relevant for risk stratification. For example, the proportion of Tregs increased with disease progression from IgM MGUS to SWM. This result is in line with a prior study reporting that WM cells are capable of inducing Treg differentiation and expansion, which in turn enables tumor growth³⁴. Therefore, immune profiling may be useful for patient risk stratification and potentially selecting patients who may benefit from early treatment for prevention of progression.

By comparing changes in immune cell proportions between patients with AWM and patients with SMM, another BM malignancy arising from antibody-producing cells, we described immune hallmarks of WM and MM; namely, types of immune dysregulation that are more disease-specific. We showed that WM exhibits a specific loss of interferon-stimulated and activated lymphocytes and a more pronounced increase in S100B⁺ CD56_{dim} NK cells and CD16⁺ Monocytes compared to SMM, while SMM exhibits a specific increase in CD4⁺ naïve T cells, a specific loss of CD14⁺ Monocytes, and a more pronounced increase in Tregs compared to AWM. We then showed that compositional differences could be used to diagnose the presence of malignancy from immune profiling alone and also differentiate between AWM and SMM, suggesting that immune profiling may have the potential to diagnose and differentiate B cell malignancies in the BM. This disease specificity in immune dysregulation extended to gene expression changes as well. For example, myeloid cells from patients with AWM showed a disease-specific upregulation of *MNDA*, a key interferon regulator, which may be related to the observed depletion of interferon-stimulated lymphocytes^{40–44}.

Probing this defect in interferon stimulation further, we demonstrated that T and NK cells from patients with AWM are hypo-responsive to interferon stimulation, a systemic defect that could be partially rescued by exogenous administration of IFN type I. This, we showed, was despite the presence of higher than normal IFN stimulation levels in CD14⁺ Monocytes from those patients, as well as normal levels of IFN- γ in PB plasma from patients with AWM. These results suggest that the observed defect could not be simply attributed to lower levels of interferon in patients with AWM, but rather may truly be cell-intrinsic. Two small prior studies demonstrated modest clinical responses in patients with WM treated with interferon, suggesting that this defect may translate into a therapeutic vulnerability in some patients^{52,53}. Using a flow cytometry-based cytotoxicity assay, we showed that NK cells from patients with AWM are indeed dysfunctional and the administration of IFN-I can help improve although not entirely correct their functionality. Collectively, these results suggest that newer formulations of IFN-I with reduced toxicity and optimized dosing could potentially have a role in WM therapy in the future, along with other treatments that potentiate NK cell functionality, although this requires further testing.

Next, we sought to characterize tumor-intrinsic mechanisms of disease progression. We defined gene expression signatures that could be used for the subclassification of WM tumors: an *ACTG1/S100A4*-high subset (corresponding to GEX-2), a *DUSP22/CD9*-high subset (corresponding to GEX-5), and a *CXCR4/AHNAK*-high subset (corresponding to GEX-6 and GEX-7), which could be further subclassified into *CXCR4/AHNAK/BCL7A*-high (corresponding to GEX-8) or low groups. The expression of surface markers CD9 and CD23 which co-segregated with some of these disease subgroups could help in the clinical translation of this classification system. The co-segregation of CD23 and *CXCR4* expression in particular has not been previously described and could be relevant for the identification of *CXCR4*-mutant patients who exhibit suboptimal responses to treatment^{3,23,62,73–76,84}. Two of the gene expression signatures (GEX-2 and GEX-7) were associated with progression risk, with higher GEX-2 activity in patients with IgM MGUS and higher GEX-7 activity in patients with overt WM. Signature GEX-7, which was marked by *CXCR4*, was shown to be active in

tumors with and without reported *CXCR4* mutations, suggesting that it may help identify patients who would otherwise not be considered at high risk of progression. In an external bulk RNA-seq cohort of patients with overt WM, signature GEX-7 was significantly correlated with the degree of BM infiltration, suggesting that it may be prognostically relevant even in the setting of symptomatic disease. Signature GEX-5, which was marked by *DUSP22* and *CD9*, was associated with a higher risk of progression and the presence of lymphadenopathy. *CD9* has been previously described to denote increased dependency of lymphoma cells on follicular dendritic cells, making disease progression beyond germinal centers less likely, which may explain the lymph node tropism of this subtype^{68,69}. A prior study observed an association between *DUSP22/CD9* expression and a plasmacytoid phenotype, suggesting that this expression program may also have morphological correlates⁸⁶. Lastly, we showed that signature GEX-5 may be more active in *CXCR4*-WT patients and therefore it may be associated with better response to treatment with Ibrutinib, although this would need to be prospectively validated^{3,62,74–76}.

In conclusion, we have comprehensively described the types of immune dysregulation observed in patients with AWM, identified immune hallmarks of disease and biomarkers of disease progression, and demonstrated systemic hypo-responsiveness of patient T and NK cells to interferon stimulation, which may result in a therapeutic vulnerability. Moreover, we uncovered significant heterogeneity within WM tumors which could be distilled into a molecular classification system, and identified tumor-intrinsic gene expression signatures associated with disease progression, which may help direct clinical stratification and therapy selection in the future.

Methods

Patient samples and processing

Participants were enrolled on one of the following studies: the PCROWD study, an observational prospective cohort study of plasma cell pre-malignancies, (IRB #14-174), the PANGAEA study (IRB #21-127), and a biobanking study of bone marrow samples from patients with plasma cell malignancies at the Dana-Farber Cancer Institute (IRB #07-150). All participants provided written informed consent prior to the collection of data and specimens. The institutional review board of the Dana-Farber Cancer Institute approved the studies in accordance with the Declaration of Helsinki. Four bone marrow aspirates (patients 1-4) were collected and processed at the National and Kapodistrian University of Athens, Athens, Greece, and shipped to the Dana-Farber on dry ice; they were sequenced at Dana-Farber under the protocol of the PANGAEA study.

Electronic medical records were reviewed independently by two people for clinical data collection. Patients with a monoclonal IgM protein in the serum and no morphological evidence of lymphoma infiltration in the bone marrow were considered to have IgM MGUS, while patients with any degree of infiltration were considered to have SWM, according to the recommendations of the 2nd International Workshop on WM^{7,88}. The patients' *MYD88* and *CXCR4* mutation status were assessed clinically using a clinical-grade deep targeted sequencing assay developed at the Dana-Farber Cancer Institute (Rapid Heme Panel) or, for *MYD88* alone, allele-specific PCR⁶⁰. Fluorescence in situ hybridization (FISH) was used clinically to detect the presence of copy number abnormalities. Patient characteristics are summarized in Table 1.

Bone marrow aspirates and peripheral blood were collected in EDTA tubes, and bone marrow mononuclear cells (BMMCs) and peripheral blood mononuclear cells (PBMCs) were isolated using ficoll separation or Red Blood Cell lysis buffer (ThermoFisher). Bone marrow mononuclear cells were then subjected to magnetic bead enrichment (Miltenyi Biotec) for CD138 and/or CD19, according to the manufacturer's instructions, and cryopreserved in Fetal Bovine Serum (FBS)

with 10% Dimethylsulfoxide (DMSO). PBMCs were cryopreserved in FBS with 10% DMSO without further selection.

Statistics and Reproducibility

We performed power analysis to determine the appropriate sample size. We estimated that approximately 20 samples/group would be required to have >90% power to detect a difference of at least 3% in cell type proportions between patients and healthy donors, assuming cell type proportions are beta-distributed and assuming a 5% frequency in healthy donors, a standard deviation of 3% in both groups, and multiple hypotheses testing correction for 20 cell types at the 0.05 significance level. No data were excluded from the analyses, unless expressly stated otherwise in the main text and figure legend (for example, removing individuals with less than 100 immune cells from compositional analyses). No randomized experiments were included in the study.

Single-cell sequencing of tumor and immune cells

We performed single-cell RNA and BCR sequencing on CD138⁺ and/or CD19⁺ tumor cells from 18 patients with IgM MGUS/SWM (IgM MGUS: $n=1$; SWM: $n=17$), and 4 patients with WM for a total number of 26 samples to molecularly characterize early-stage tumors in patients with IgM MGUS/SWM (Table 1). For 1 patient, both CD138⁺ and CD19⁺ cells were sequenced; for another, two aliquots were thawed and for each aliquot, both CD138⁺ and CD19⁺ cells were sequenced. One patient with AWM (Pt5) had a serial sample collected at progression 5 years after the original sample was drawn. We also performed single-cell RNA sequencing on CD138⁺ and/or CD19⁺ immune cells from 25 patients with IgM MGUS/SWM (IgM MGUS: $n=6$; SWM: $n=19$) and 3 patients with WM for a total number of 29 samples (one patient had both a CD138⁺ and a CD19⁺ sample) to comprehensively characterize alterations in the BM immune microenvironment across stages of disease progression (Table 1).

Cells were thawed in a 37 °C water bath. Subsequently, they were centrifuged at 330 g for 5 min and washed twice with an ice-cold 0.04% Ultrapure Bovine Serum Albumin (BSA)/Phosphate-Buffered Saline (PBS) wash buffer, before being loaded onto a Chromium Controller (10X Genomics) for single cell encapsulation. Libraries were prepared using the Chromium Next GEM Single Cell 5' Reagent Kit v2 (Dual Index), the Chromium Single Cell Human BCR Amplification Kit, and Library Construction Kits (10X Genomics), according to the manufacturer's instructions. Libraries were sequenced on a NovaSeq 6000 S4 flow cell at the Genomics Platform of the Broad Institute of MIT and Harvard (Cambridge, MA).

Single-cell RNA sequencing data processing and analysis

Cell Ranger (v6.0.1) was used to demultiplex FASTQ files and produce count matrices and R (v4.1.3) and Seurat (v4.1.0) were used for downstream analyses^{89,90}. We performed ambient RNA correction with SoupX (v1.5.2); doublet detection with Scrublet (v0.2.3), scDblFinder (v1.8.0), and SCDS (v1.10.0); and normalization with Scraper (v1.22.1)^{91–95}. Tumor and immune fraction samples from all 30 patients with AWM/WM were processed together and integrated with in-house data from BM immune cell samples from healthy donors (HD, $n=23$) and patients with Smoldering Multiple Myeloma (SMM; $n=26$)^{36,37}. A total of 292 stroma cells were identified and removed from all analyses. The total number of samples with at least 100 immune cells annotated (excluding normal B cells), who were considered for compositional analyses, was 70 (HD: 18; IgM MGUS: 6; SWM: 19; WM: 3; SMM: 24), including one patient who was sampled at both the SWM and WM stages. A total of 10 samples (HD: 5, SWM: 2; WM: 1; SMM: 2) had fewer than 100 immune cells annotated and were excluded from analyses.

Following the identification of malignant cells, malignant and immune cells were processed separately. For immune cells, integration was performed using Harmony (v0.1.0) and correcting for sample ID⁹⁶.

Droplets were deemed to be doublets when at least 2 out of three methods classified them as such and they were only removed from consideration when they clustered together and co-expressed markers of multiple cell types (heterotypic doublets). Droplets containing dying cells with more than 15% mitochondrial gene expression were removed before clustering. Clusters with higher mitochondrial and ribosomal gene expression that clearly separated from well-annotated clusters and lacked interpretable expression markers were removed downstream. Cells were annotated based on a list of established expression markers, as well as cluster-specific markers obtained through differential expression analysis, as previously described³⁶.

Cell type proportions were computed using all cells as denominator and compared between groups using two-sided Wilcoxon's rank-sum tests (only considering samples with at least 100 annotated immune cells). P -values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg approach⁹⁷.

Genes with more than 3 counts in at least 10 cells were considered for differential expression analysis. For each cell subtype, cells of one group (for example, AWM) were compared to cells of the other (for example, HD) using two-sided Wilcoxon's rank-sum tests and p -values were corrected with the Benjamini-Hochberg approach⁹⁷. To systematically compare gene expression changes in AWM and SMM by cell type, cell type-specific results (for example, for myeloid cells) were aggregated by computing the mean log₂ fold-change and q -value per gene across subtype-specific comparisons (for example, CD16⁺ Monocytes, S100A⁺ CD14⁺ Monocytes, etc.).

Immune cell proportion classifier

For the purposes of this analysis, we discarded progenitor cells and focused on individuals with at least 100 immune cells excluding progenitor cells (HD: $n=17$; SMM: $n=24$; AWM: $n=25$). Progenitor cells were removed to control for the effect of concurrent clonal hematopoiesis which is frequently encountered in patients with AWM and SMM and may lead to changes in the proportion of different progenitor populations^{35,98}. We selected 17 features (i.e., cell types) using the rank-based approach described by Golub et al. and taking the union of the top 10 features for each of the two comparisons (SMM vs HD, AWM vs HD)⁴⁵. We then split our cohort into five subsets and performed 5-fold cross-validation, each time training an SVM classifier on the training subsets and testing it on the testing subset. For each sample, the SVM classifier determined the diagnosis (i.e., HD, SMM, or AWM). The function svm() from the R package e1071 (v1.7.11) was used to train the classifier with cost set to 0.1 and a linear kernel.

Interferon and activation signature scoring

To measure interferon signaling and general immune cell activation, we used the function AddModuleScore() of the R package Seurat (v4.1.0) and scored cells for these gene sets, respectively: (i) *ISG15*, *ISG20*, *IFI6*, *IFI27*, *IFI44L*, *MX1*, *STAT1*, *MX2*, *OAS1*, *OASL*, *EIF2AK2*, *IFIT1*, *IFIT2*, *IFIT3*, and (ii) *JUN*, *JUNB*, *JUND*, *FOS*, *FOSB*, *CD69*, *NR4A2*, *DUSP1*, *TNFAIP3*, *TSC22D3*, *PPP1R15A*, *FAM177A1*, *ZFP36*, *ZFP36L2*, *FTH1*, *KLF6*, *BTG1*, *ZNF331*, *EIF1*⁹⁰. Signature activity was compared between groups of cells using Wilcoxon's rank-sum test and p -values were corrected using the Benjamini-Hochberg approach⁹⁷.

IFN-γ expression post-stimulation with common viral epitopes

Cryopreserved PBMCs from 3 patients with AWM (IgM MGUS: $n=2$; SWM: $n=1$), 10 patients with SMM, and 10 HD were thawed in a 37 °C water bath, spun down, and resuspended in CTL-Test Medium (Cellular Technology Limited, Shaker Hts., OH). For every individual, 400,000 cells were seeded per well in duplicate in 100 uL of CTL-Test media. For antigen stimulation, we used a peptide pool consisting of 124 class I-restricted epitopes from Cytomegalovirus (30 peptides), Epstein Barr Virus (59 peptides), Human Respiratory Syncytial Virus (11 peptides), and Influenza A Virus (24 peptides) (CTL-CERI-300, Cellular

Technology Limited) at a final concentration of 1 µg/ml. Negative control wells ($n=2$) were supplemented with 0.3 µL DMSO (0.3%), instead. The Human IFN-γ Single-Color Enzymatic ELISPOT assay kit (hIFNg-2M/2, Cellular Technology Limited, Shaker Hts., OH) was used to assess the expression of IFN-γ post-stimulation. The average number of spots across duplicate test wells was divided by that in the negative control wells to normalize estimates. Antigen-specific responses were expressed as ratios (test/control) and compared between groups using Wilcoxon's rank-sum tests. *P*-values were corrected using the Benjamini-Hochberg approach⁹⁷.

IFN type I stimulation of BMMCs and PBMCs

Cryopreserved CD138⁺ BMMCs from 3 patients with AWM and 3 HD as well as cryopreserved PBMCs from the same 3 patients with AWM were cultured with or without 1000 U/mL of universal Type I IFN (R&D systems) for 16 h in RPMI 1640 medium (Gibco) containing 20% FBS. Following stimulation, cells were centrifuged at 330 g for 5 min and washed with 0.04% Ultrapure BSA/PBS, before being loaded on the Chromium Controller (10X Genomics) for single cell encapsulation. Libraries were prepared using the Chromium Next GEM Single Cell 5' Reagent Kit v2 (Dual Index) and Library Construction Kits (10X Genomics), according to the manufacturer's instructions, and sequenced on a NovaSeq 6000 S4 flow cell at the Genomics Platform of the Broad Institute of MIT and Harvard (Cambridge, MA). The data was analyzed as described above.

Measurement of IFN-γ levels in peripheral blood plasma

Peripheral blood plasma was collected from 5 patients with IgM MGUS, 15 patients with non-IgM MGUS, 20 patients with SMM, and 13 HD. Peripheral blood was collected in EDTA tubes and subjected to centrifugation at 1200 g for 10 min. Plasma was collected and subjected to a second round of centrifugation at 2000 g for 20 min to remove remaining cells and debris. Plasma was then aliquoted in cryotubes and stored at -80 °C until use. Proximity Extension Assay technology by Olink Proteomics (Uppsala, Sweden) was used to measure the levels of IFN-γ in plasma samples and cross-group comparisons were conducted with Wilcoxon's rank-sum tests. *P*-values were corrected using the Benjamini-Hochberg approach⁹⁷.

Flow cytometry-based NK cell cytotoxicity assay

An NK cell cytotoxicity assay was performed as previously described in ref. 99. Briefly, NK cells were isolated from cryopreserved PBMCs from 5 patients with AWM and 4 HD using EasySep Human NK Cell Isolation kit (StemCell Technologies) and cultured overnight in RPMI-1640 medium (Gibco) containing 10% FBS, 1X penicillin/streptomycin, and 1 ng/mL IL-15 (PeproTech). K562 cells expressing GFP were kindly provided by Dr. Eric Smith (Dana-Farber Cancer Institute, Boston, MA) and cultured in RPMI-1640 medium containing 10% FBS and 1X penicillin/streptomycin. NK cells and K562 cells expressing GFP were co-cultured at an effector/target (E:T) ratio of 3:1 with or without 1000 U/mL of universal Type I IFN (R&D systems) for 4 h in RPMI-1640 medium containing 10% FBS, 1X penicillin/streptomycin, and 1 ng/mL IL-15. Following co-culture, cells were stained with PE-annexin V (Biolegend) and 7-AAD (BD Biosciences) for 15 min at room temperature. Cells were then acquired using a BD LSR Fortessa X-20 instrument at the Flow Cytometry Core of the Dana-Farber Cancer Institute. The data was collected using BD FACS Diva software (v. 9.2.0) and analyzed using FlowJo (v.10.8.1). Cells were first gated on GFP to identify GFP⁺ K562 cells; doublets were excluded using FSC-H vs FSC-W and then SSC-H vs SSC-W; killed target cells were then identified based on positive staining for both PE-Annexin V and 7-AAD.

MYD88 L265P and CXCR4 mutation detection assays

Assessment of MYD88 L265P and CXCR4 mutation status was performed using DNA extracted from CD138⁺ and/or CD19⁺ bone marrow

cells from 18 patients with AWM and available samples with the Monarch Genomic DNA Purification Kit (New England Biolabs). Quantitative PCR (qPCR) was performed using Power SYBR Green PCR Master Mix according to the manufacturer's instructions on the ABI Prism 7500 Sequence Detection System (Applied Biosystems), as previously described^{100,101}. Briefly, the PCR reaction was performed in a final volume of 25 µL with 25 nM of each primer and 50 ng of DNA. Thermal cycling conditions were as follows: 10 min at 95 °C, followed by 40 cycles of 95 °C for 15 s and 60 °C for 60 s. For MYD88 L265P, the mutant-specific reverse primer (5'-CCTTGTACTTGATGGGAACG-3'), the wild-type-specific reverse primer (5'-TGGTGTAGTCGACAGCAGTGA-3'), and the common forward primer (5'-AATGTGTGC-CAGGGGTACTTAG-3') were used. For CXCR4^{S338X} mutations, the C > G mutation-specific reverse primer (5'-AGACTCAGCTCAGTGGAAA-CAGTTC-3'), the C > A mutation-specific reverse primer (5'-AGACTCAGCTCAGTGGAAACAGGT-3'), the common forward primer (5'-TTCCACTGTTGTCTGAACCCATC-3'), the reference forward primer (5'-ACTACATTGGGATCAGCATCGACT C-3') and the reference reverse primer (5'-TGAAGACTCAGCTCAGTGGAAACAG-3') were used. The standard curve for MYD88 L265P and CXCR4 mutations were generated by a serial dilution of the mutant DNA with the wild-type DNA (50%, 10%, 2%, 0.4%, 0.08%, and wild-type), respectively. To detect CXCR4 mutations at the C-terminal domain, Sanger sequencing was performed as previously described¹⁰⁰. Briefly, the forward PCR primer (5'-ATGGGGAGGAGTTGTAGGATTCTAC-3') and reverse PCR primer (5'-TTGGCCACAGGTCCTGCCTAGACA-3') were designed to amplify the CXCR4 open reading frame. Amplified PCR products were isolated using the QIA quick gel extraction kit (Qiagen) and sequenced using both forward and reverse PCR primers and an additional sequencing primer (5'-GCTGCCTTACTACATTGGGATCAGC-3').

Single-cell BCR sequencing data processing

B cells and plasma cells were first identified based on the expression of key lineage markers (B cells: *CD19*, *PAX5*, *MS4A1*, *CD79A*; Plasma cells: *SDC1*, *TNFRSF17*, *IRF4*, *XBPI*, *SLAMF7*, *CD38*). Cell barcodes that were determined to correspond to B cells and plasma cells were considered for downstream analysis. All V(D)J contigs detected by CellRanger vdj (v6.0.1) were processed to identify a single heavy and light chain per cell barcode, based on UMI support⁸⁹. For each patient, clonotypic data from multiple samples or aliquots were combined and harmonized clonotypes were identified anew based on the frequency of unique Complementarity-Determining Region 3 (CDR3) amino-acid (AA) sequences.

Identification of malignant cells and tumor phenotyping

To identify malignant cells, B cells and plasma cells from each patient were considered separately. For each patient, B cell receptor clonotypes were sorted by their frequency, and clonotypes whose frequency was at least double that of the next clonotype in order and which had at least 30 cells, were nominated as expanded. Expanded clonotypes which clustered separately from normal B cells and whose immunoglobulin isotype matched the one detected clinically by immunofixation were considered malignant. Cells residing in the same clusters as the malignant clonotypes were considered malignant. A panel of established markers was then used to determine the type of the lymphoid clone. For example, WM clones were identified based on the expression of *IGHM*, *CD79B*, *CD27*, *TNFRSF13B*, *ITM2C*, *JCHAIN*, and *CD1C*. Concurrent Monoclonal B cell Lymphocytosis (MBL)/Chronic Lymphocytic Leukemia (CLL) clones were identified based on the expression of *CD5*, *FCER2* (encoding CD23), *SPN* (encoding CD43), and *LEFI*.

Inference of Copy Number Variants from scRNA-seq data

Copy number variants (CNVs) were inferred from scRNA-seq data using Numbat (v1.1.0)⁶³. Allelic data was collected from the entire

positive fraction, which contained both tumor and normal cells, to ensure that the frequency of CNVs and particularly, clonal deletions, would not negatively impact genotyping and phasing. A custom panel of 1.2 K healthy donor plasma cells was used as expression reference to correct for cell type-specific expression patterns, such as immunoglobulin expression, which otherwise may result in artifactual CNV inference on chromosomes 2, 14, and 22. Ultimately, CNVs were inferred on tumor cells only to increase sensitivity.

De novo extraction of gene expression signatures from tumor cells

SignatureAnalyzer-GPU (v0.0.8) was used for signature discovery on a matrix of WM tumor cells and the top 2000 highly variable genes, excluding immunoglobulin, ribosomal, and mitochondrial genes^{36,102–104}. The tool was run 30 times directly on the raw count matrix with a Poisson objective and exponential (L1) priors for both the H and the W matrices. The run that resulted in a K (i.e., number of signatures) equal to the mode of the K distribution and had the lowest objective was selected for downstream analysis. Gene expression signature markers were nominated by (i) multiplying the W matrix by the sum of signature activity across all cells in the H matrix, (ii) calculating the fraction of each signature's activity for each gene (matrix F) and (iii) ranking genes based on the product of W (i.e., how strongly each gene contributes to the signature) and F (i.e., how strongly each signature contributes to the gene)³⁶. A total of 8 gene expression signatures (GEX) were extracted. Signatures GEX-1 and GEX-4 captured general immune cell activation through the AP-1 module and monocyte contamination, respectively, and were removed from downstream analyses.

Validation of gene expression signatures in an external GEP dataset

To validate our gene expression signatures, we used an external GEP dataset of 13 patients with IgM MGUS and 36 patients with overt WM (GSE171739)²⁷. The dataset was downloaded from the Gene Expression Omnibus (GEO) using the getGEO() function from the GEOquery (v2.68.0) R package¹⁰⁵. CD19⁺ patient samples (IgM MGUS: n = 13; WM: n = 36) were selected for downstream analysis and gene-level data was generated by averaging across each gene's probes. Normalized expression values for the 30 markers of our six gene expression signatures (5 markers per signature) were z-scored and a Euclidean distance matrix was constructed for agglomerative hierarchical clustering with complete linkage. The number of clusters was set to 6, one per signature, and the final number of subtypes was determined to be 4 based on manual curation for ease of classification. Samples were then scored for each signature by taking the mean of the signature's markers' expression levels and each signature's activity was compared between patients with IgM MGUS and patients with WM using two-sided Wilcoxon's rank-sum tests. *P*-values were adjusted using the Benjamini-Hochberg approach⁹⁷.

Subsequently, a list of 101 genes that were shown to be consistently upregulated in more than 30% of tumors in our single-cell RNA seq dataset were tested for differential expression between patients with IgM MGUS and patients with overt WM in GSE171739 using two-sided Wilcoxon's rank-sum tests²⁷. *P*-values were adjusted using the Benjamini-Hochberg approach and genes with a *q*-value < 0.05 (*n* = 53) were selected for visualization⁹⁷.

Validation of gene expression signatures in an external bulk RNA-seq dataset

To further validate our gene expression signatures, we used an external bulk RNA-seq dataset of 52 patients with *MYD88*-mutant WM and available clinical and genotypic data on *CXCR4* mutations⁶¹. Normalized expression data and associated metadata was obtained following communication with the authors. Samples were scored for

signatures GEX-2, GEX-5, and GEX-7 by taking the mean of the signature's markers' expression levels. Signature activity was compared between patients of different genotypic status, patients with or without LAD, and patients with or without splenomegaly using Wilcoxon's rank-sum tests; signature activity was compared to BM infiltration and serum IgM levels using Pearson's correlation tests. *P*-values were corrected using the Benjamini-Hochberg approach⁹⁷.

Reporting summary

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

Data availability

Source data are provided with this paper. Raw and processed single-cell RNA and BCR sequencing data generated for this study are deposited in dbGaP (phs003787.v1.p1; https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs003787.v1.p1). Raw single-cell RNA sequencing data generated in Zavidij et al. are deposited in dbGaP (phs001323.v3.p1; https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001323.v3.p1)³⁷. Raw single-cell RNA sequencing data generated in Sklavenitis-Pistofidis et al. are deposited in dbGaP (phs002476.v3.p1; https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002476.v3.p1)³⁶. Normalized bulk RNA-seq data from Hunter et al. was obtained directly from the authors and is not deposited in a public repository; signature scores generated on this cohort for the purposes of this manuscript are shared as Source Data⁶¹. Gene expression profiling data from Trojani et al. was downloaded from GEO (GSE171739; <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE171739>)²⁷. The remaining data are available within the Article, Supplementary Information or Source Data file. Source data are provided with this paper.

Code availability

Code used for downstream analysis is available on Github: <https://github.com/romanos-sp/Nat-Comm-2025-WM-scRNA-seq> or as Supplementary Software.

References

- Wang, H. et al. Temporal and geographic variations of Waldenstrom macroglobulinemia incidence: a large population-based study. *Cancer* **118**, 3793–3800 (2012).
- Garcia-Sanz, R. et al. New developments in the diagnosis and characterization of Waldenstrom's macroglobulinemia. *Expert Rev. Hematol.* **16**, 835–847 (2023).
- Treon, S. P. et al. Genomic Landscape of Waldenstrom Macroglobulinemia and Its Impact on Treatment Strategies. *J. Clin. Oncol.* **38**, 1198–1208 (2020).
- Kyle, R. A. et al. IgM monoclonal gammopathy of undetermined significance and smoldering Waldenstrom's macroglobulinemia. *Clin. Lymphoma Myeloma* **9**, 17–18 (2009).
- Kyle, R. A. et al. Long-term follow-up of IgM monoclonal gammopathy of undetermined significance. *Blood* **102**, 3759–3764 (2003).
- Kyle, R. A. et al. Progression in smoldering Waldenstrom macroglobulinemia: long-term results. *Blood* **119**, 4462–4466 (2012).
- Owen, R. G. et al. Clinicopathological definition of Waldenstrom's macroglobulinemia: consensus panel recommendations from the Second International Workshop on Waldenstrom's Macroglobulinemia. *Semin Oncol.* **30**, 110–115 (2003).
- Gertz, M. A. Waldenstrom macroglobulinemia: 2023 update on diagnosis, risk stratification, and management. *Am. J. Hematol.* **98**, 348–358 (2023).
- Ansell, S. M. et al. Diagnosis and management of Waldenstrom macroglobulinemia: Mayo stratification of macroglobulinemia

- and risk-adapted therapy (mSMART) guidelines. *Mayo Clin. Proc.* **85**, 824–833 (2010).
10. Bustoros, M. et al. Progression Risk Stratification of Asymptomatic Waldenstrom Macroglobulinemia. *J. Clin. Oncol.* **37**, 1403–1411 (2019).
 11. Merli, M. et al. Validation and further potentialities of the novel AWM score for progression risk stratification in patients with asymptomatic Waldenstrom macroglobulinemia. *Leuk. Lymphoma* **61**, 987–989 (2020).
 12. Kyle, R. A. et al. Long-Term Follow-up of Monoclonal Gammopathy of Undetermined Significance. *N. Engl. J. Med.* **378**, 241–249 (2018).
 13. Moreno, D. F. et al. Defining an Ultra-Low Risk Group in Asymptomatic IgM Monoclonal Gammopathy. *Cancers (Basel)* **13**, 2055 (2021).
 14. Zanwar, S. et al. Disease outcomes and biomarkers of progression in smouldering Waldenstrom macroglobulinaemia. *Br. J. Haematol.* **195**, 210–216 (2021).
 15. Varettoni, M. et al. A risk-stratification model based on the initial concentration of the serum monoclonal protein and MYD88 mutation status identifies a subset of patients with IgM monoclonal gammopathy of undetermined significance at high risk of progression to Waldenstrom macroglobulinaemia or other lymphoproliferative disorders. *Br. J. Haematol.* **187**, 441–446 (2019).
 16. Zanwar, S. et al. Impact of MYD88(L265P) mutation status on histological transformation of Waldenstrom Macroglobulinemia. *Am. J. Hematol.* **95**, 274–281 (2020).
 17. Bustoros, M. et al. Biomarkers of Progression and Risk Stratification in Asymptomatic Waldenstrom Macroglobulinemia. *Hematol. Oncol. Clin. North Am.* **37**, e1–e13 (2023).
 18. Treon, S. P. et al. MYD88 L265P somatic mutation in Waldenstrom's macroglobulinemia. *N. Engl. J. Med.* **367**, 826–833 (2012).
 19. Moreno, D. F. et al. Prognostic impact of MYD88 and CXCR4 mutations assessed by droplet digital polymerase chain reaction in IgM monoclonal gammopathy of undetermined significance and smouldering Waldenstrom macroglobulinaemia. *Br. J. Haematol.* **200**, 187–196 (2023).
 20. Varettoni, M. et al. Pattern of somatic mutations in patients with Waldenstrom macroglobulinemia or IgM monoclonal gammopathy of undetermined significance. *Haematologica* **102**, 2077–2085 (2017).
 21. Hunter, Z. R. et al. The genomic landscape of Waldenstrom macroglobulinemia is characterized by highly recurring MYD88 and WHIM-like CXCR4 mutations, and small somatic deletions associated with B-cell lymphomagenesis. *Blood* **123**, 1637–1646 (2014).
 22. Roccaro, A. M. et al. C1013G/CXCR4 acts as a driver mutation of tumor progression and modulator of drug resistance in lymphoplasmacytic lymphoma. *Blood* **123**, 4120–4131 (2014).
 23. Sklavenitis-Pistofidis, R. et al. Bortezomib overcomes the negative impact of CXCR4 mutations on survival of Waldenstrom macroglobulinemia patients. *Blood* **132**, 2608–2612 (2018).
 24. Garcia-Sanz, R. et al. 6q deletion in Waldenstrom macroglobulinaemia negatively affects time to transformation and survival. *Br. J. Haematol.* **192**, 843–852 (2021).
 25. Maclachlan, K. H. et al. Waldenstrom macroglobulinemia whole genome reveals prolonged germinal center activity and late copy number aberrations. *Blood Adv.* **7**, 971–981 (2023).
 26. Guerrero, M. L. et al. MYD88 mutated and wild-type Waldenstrom's Macroglobulinemia: characterization of chromosome 6q gene losses and their mutual exclusivity with mutations in CXCR4. *Haematologica* **103**, e408–e411 (2018).
 27. Trojani, A. et al. Identification of a Candidate Gene Set Signature for the Risk of Progression in IgM MGUS to Smoldering/Symptomatic Waldenstrom Macroglobulinemia (WM) by a Comparative Transcriptome Analysis of B Cells and Plasma Cells. *Cancers (Basel)* **13**, 1837 (2021).
 28. Paiva, B. et al. The cellular origin and malignant transformation of Waldenstrom macroglobulinemia. *Blood* **125**, 2370–2380 (2015).
 29. Jalali, S. & Ansell, S. M. The Bone Marrow Microenvironment in Waldenstrom Macroglobulinemia. *Hematol. Oncol. Clin. North Am.* **32**, 777–786 (2018).
 30. Jalali, S. et al. Soluble PD-1 ligands regulate T-cell function in Waldenstrom macroglobulinemia. *Blood Adv.* **2**, 1985–1997 (2018).
 31. Hodge, L. S. et al. IL-21 in the bone marrow microenvironment contributes to IgM secretion and proliferation of malignant cells in Waldenstrom macroglobulinemia. *Blood* **120**, 3774–3782 (2012).
 32. Cholujova, D. et al. Dysfunctions of innate and adaptive immune tumor microenvironment in Waldenstrom macroglobulinemia. *Int. J. Cancer* **152**, 1947–1963 (2023).
 33. Kaushal, A. et al. Aberrant Extrafollicular B Cells, Immune Dysfunction, Myeloid Inflammation, and MyD88-Mutant Progenitors Precede Waldenstrom Macroglobulinemia. *Blood Cancer Discov.* **2**, 600–615 (2021).
 34. Sacco, A. et al. Targeting the immune microenvironment in Waldenstrom macroglobulinemia via halting the CD40/CD40-ligand axis. *Blood* **141**, 2615–2628 (2023).
 35. Tahri, S. et al. Clonal hematopoiesis is associated with increased risk of progression of asymptomatic Waldenstrom macroglobulinemia. *Blood Adv.* **6**, 2230–2235 (2022).
 36. 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).
 37. Zavidij, O. et al. Single-cell RNA sequencing reveals compromised immune microenvironment in precursor stages of multiple myeloma. *Nat. Cancer* **1**, 493–506 (2020).
 38. Konishi, Y. et al. Attenuated response to SARS-CoV-2 vaccine in patients with asymptomatic precursor stages of multiple myeloma and Waldenstrom macroglobulinemia. *Cancer Cell* **40**, 6–8 (2022).
 39. Jauch-Speer, S. L. et al. C/EBPdelta-induced epigenetic changes control the dynamic gene transcription of S100a8 and S100a9. *Elife* **11**, e75594 (2022).
 40. Briggs, J. A. et al. Cloning and expression of the human myeloid cell nuclear differentiation antigen: regulation by interferon alpha. *J. Cell Biochem* **49**, 82–92 (1992).
 41. Briggs, R. et al. Interferon alpha selectively affects expression of the human myeloid cell nuclear differentiation antigen in late stage cells in the monocytic but not the granulocytic lineage. *J. Cell Biochem* **54**, 198–206 (1994).
 42. Briggs, R. C. et al. The human myeloid cell nuclear differentiation antigen gene is one of at least two related interferon-inducible genes located on chromosome 1q that are expressed specifically in hematopoietic cells. *Blood* **83**, 2153–2162 (1994).
 43. Burrus, G. R., Briggs, J. A. & Briggs, R. C. Characterization of the human myeloid cell nuclear differentiation antigen: relationship to interferon-inducible proteins. *J. Cell Biochem* **48**, 190–202 (1992).
 44. Gu, L. et al. Myeloid cell nuclear differentiation antigen controls the pathogen-stimulated type I interferon cascade in human monocytes by transcriptional regulation of IRF7. *Nat. Commun.* **13**, 14 (2022).
 45. Golub, T. R. et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
 46. El-Sherbiny, Y. M. et al. A novel two-score system for interferon status segregates autoimmune diseases and correlates with clinical features. *Sci. Rep.* **8**, 5793 (2018).

47. Oke, V. et al. High levels of circulating interferons type I, type II and type III associate with distinct clinical features of active systemic lupus erythematosus. *Arthritis Res. Ther.* **21**, 107 (2019).
48. Yerram, K. V. et al. Serum interferon-alpha predicts in-hospital mortality in patients hospitalised with acute severe lupus. *Lupus Sci. Med.* **10**, e000933 (2023).
49. Contoli, M. et al. Blood Interferon-alpha Levels and Severity, Outcomes, and Inflammatory Profiles in Hospitalized COVID-19 Patients. *Front Immunol.* **12**, 648004 (2021).
50. Cao, Y. et al. The cytokine/chemokine pattern in the bone marrow environment of multiple myeloma patients. *Exp. Hematol.* **38**, 860–867 (2010).
51. Kawano, Y. et al. Blocking IFNAR1 inhibits multiple myeloma-driven Treg expansion and immunosuppression. *J. Clin. Invest.* **128**, 2487–2499 (2018).
52. Rotoli, B. et al. A phase II trial on alpha-interferon (alpha IFN) effect in patients with monoclonal IgM gammopathy. *Leuk. Lymphoma* **13**, 463–469 (1994).
53. Legouffe, E. et al. Treatment of Waldenstrom's macroglobulinemia with very low doses of alpha interferon. *Leuk. Lymphoma* **19**, 337–42, (1995).
54. Wang, Q. et al. Biclinal lymphoplasmacytic lymphoma/Waldenstrom macroglobulinemia associated with POEMS syndrome: A case report and literature review. *Oncol. Lett.* **25**, 97 (2023).
55. Kyle, R. A., Robinson, R. A. & Katzmann, J. A. The clinical aspects of biclinal gammopathies. Review of 57 cases. *Am. J. Med.* **71**, 999–1008 (1981).
56. Sharma, S. et al. Demystifying Biclinal Gammopathy: A Pathologist's Perspective. *Lab Med.* **50**, 357–363 (2019).
57. Tschumper, R. C. et al. Molecular analysis of immunoglobulin genes reveals frequent clonal relatedness in double monoclonal gammopathies. *Blood Cancer J.* **3**, e112 (2013).
58. Lemal, R. et al. Mast cell density and its clinical relevance in Waldenstrom's macroglobulinemia. *E. J. Haem.* **3**, 371–378 (2022).
59. Tournilhac, O. et al. Mast cells in Waldenstrom's macroglobulinemia support lymphoplasmacytic cell growth through CD154/CD40 signaling. *Ann. Oncol.* **17**, 1275–1282 (2006).
60. Kluk, M. J. et al. Validation and Implementation of a Custom Next-Generation Sequencing Clinical Assay for Hematologic Malignancies. *J. Mol. Diagn.* **18**, 507–515 (2016).
61. Hunter, Z. R. et al. Transcriptome sequencing reveals a profile that corresponds to genomic variants in Waldenstrom macroglobulinemia. *Blood* **128**, 827–838 (2016).
62. Kaiser, L. M. et al. CXCR4 in Waldenstrom's Macroglobulinemia: chances and challenges. *Leukemia* **35**, 333–345 (2021).
63. Gao, T. et al. Haplotype-aware analysis of somatic copy number variations from single-cell transcriptomes. *Nat. Biotechnol.* **41**, 417–426 (2023).
64. Hossain, S. et al. The RASSF6 Tumor Suppressor Protein Regulates Apoptosis and Cell Cycle Progression via Retinoblastoma Protein. *Mol. Cell Biol.* **38**, e00046–18 (2018).
65. Davis, T. A., Loos, B. & Engelbrecht, A. M. AHNK: the giant jack of all trades. *Cell Signal* **26**, 2683–2693 (2014).
66. Hapgood, G. et al. Identification of high-risk DUSP22-rearranged ALK-negative anaplastic large cell lymphoma. *Br. J. Haematol.* **186**, e28–e31 (2019).
67. Onaindia, A. et al. DUSP22-rearranged anaplastic lymphomas are characterized by specific morphological features and a lack of cytotoxic and JAK/STAT surrogate markers. *Haematologica* **104**, e158–e162 (2019).
68. Yoon, S. O. et al. CD9 may contribute to the survival of human germinal center B cells by facilitating the interaction with follicular dendritic cells. *FEBS Open Bio* **4**, 370–6 (2014).
69. Yoon, S. O. et al. Down-regulation of CD9 expression and its correlation to tumor progression in B lymphomas. *Am. J. Pathol.* **177**, 377–386 (2010).
70. Balinas-Gavira, C. et al. Frequent mutations in the amino-terminal domain of BCL7A impair its tumor suppressor role in DLBCL. *Leukemia* **34**, 2722–2735 (2020).
71. Chakraborty CM, E. et al. Disruption of the m-SWI/SNF Complex Mediated By Recurrent Non-Coding Mutations in BCL7A Induces Tumor Cell Proliferation in Multiple Myeloma. *Blood* **136**, 40 (2020).
72. Caers, J. et al. Thymosin beta4 has tumor suppressive effects and its decreased expression results in poor prognosis and decreased survival in multiple myeloma. *Haematologica* **95**, 163–167 (2010).
73. Hunter, Z. R. et al. CD5, CD10, and CD23 expression in Waldenstrom's macroglobulinemia. *Clin. Lymphoma* **5**, 246–249 (2005).
74. Cao, Y. et al. The WHIM-like CXCR4(S338X) somatic mutation activates AKT and ERK, and promotes resistance to ibrutinib and other agents used in the treatment of Waldenstrom's Macroglobulinemia. *Leukemia* **29**, 169–176 (2015).
75. Cao, Y. et al. CXCR4 WHIM-like frameshift and nonsense mutations promote ibrutinib resistance but do not supplant MYD88(L265P) -directed survival signalling in Waldenstrom macroglobulinaemia cells. *Br. J. Haematol.* **168**, 701–707 (2015).
76. Castillo, J. J. et al. CXCR4 mutation subtypes impact response and survival outcomes in patients with Waldenstrom macroglobulinaemia treated with ibrutinib. *Br. J. Haematol.* **187**, 356–363 (2019).
77. Gertz, M. A. Waldenstrom macroglobulinemia: 2021 update on diagnosis, risk stratification, and management. *Am. J. Hematol.* **96**, 258–269 (2021).
78. Bagratuni, T. et al. Single-cell analysis of MYD88(L265P) and MYD88(WT) Waldenstrom macroglobulinemia patients. *Hemisphere* **8**, e27 (2024).
79. Owen, R. G., Hillmen, P. & Rawstron, A. C. CD52 expression in Waldenstrom's macroglobulinemia: implications for alemtuzumab therapy and response assessment. *Clin. Lymphoma* **5**, 278–81 (2005).
80. Treon, S. P. et al. Long-term follow-up of symptomatic patients with lymphoplasmacytic lymphoma/Waldenstrom macroglobulinemia treated with the anti-CD52 monoclonal antibody alemtuzumab. *Blood* **118**, 276–281 (2011).
81. Chakraborty, M. G., Duncavage, Z. J., Roundy, E. J., Dong, N. & Schuettelpelz, Q. The Tetraspanin CD53 Promotes CXCR Signaling to Facilitate B Cell Migration. *Blood* **140**, 8345–8346 (2022).
82. Shaffer, A. L. et al. Overcoming Acquired Epigenetic Resistance to BTK Inhibitors. *Blood Cancer Discov.* **2**, 630–647 (2021).
83. Wu, W. et al. Inhibition of B-cell receptor signaling disrupts cell adhesion in mantle cell lymphoma via RAC2. *Blood Adv.* **5**, 185–197 (2021).
84. Treon, S. P. et al. Somatic mutations in MYD88 and CXCR4 are determinants of clinical presentation and overall survival in Waldenstrom macroglobulinemia. *Blood* **123**, 2791–2796 (2014).
85. Treon, S. P., Xu, L. & Hunter, Z. MYD88 Mutations and Response to Ibrutinib in Waldenstrom's Macroglobulinemia. *N. Engl. J. Med.* **373**, 584–586 (2015).
86. Sun, H. et al. Single-cell profiles reveal tumor cell heterogeneity and immunosuppressive microenvironment in Waldenstrom macroglobulinemia. *J. Transl. Med.* **20**, 576 (2022).
87. Qiu, Y. et al. Single-cell transcriptome analysis reveals stem cell-like subsets in the progression of Waldenstrom's macroglobulinemia. *Exp. Hematol. Oncol.* **12**, 18 (2023).
88. Dogliotti, I. et al. Diagnostics in Waldenstrom's macroglobulinemia: a consensus statement of the European Consortium for Waldenstrom's Macroglobulinemia. *Leukemia* **37**, 388–395 (2023).

89. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
90. Satija, R. et al. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
91. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, giae151 (2020).
92. 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).
93. Germain, P. L. et al. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res.* **10**, 979 (2021).
94. Bais, A. S. & Kostka, D. scds: computational annotation of doublets in single-cell RNA sequencing data. *Bioinformatics* **36**, 1150–1158 (2020).
95. Lun, A. T., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res.* **5**, 2122 (2016).
96. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
97. Benjamini YH, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc.: Ser. B (Methodol.)* **57**, 289–300 (1995).
98. Mouhieddine, T. H. et al. Clonal hematopoiesis is associated with adverse outcomes in multiple myeloma patients undergoing transplant. *Nat. Commun.* **11**, 2996 (2020).
99. Shapiro, R. M. et al. Expansion, persistence, and efficacy of donor memory-like NK cells infused for posttransplant relapse. *J. Clin. Invest* **132**, e154334 (2022).
100. Xu, L. et al. Clonal architecture of CXCR4 WHIM-like mutations in Waldenstrom Macroglobulinaemia. *Br. J. Haematol.* **172**, 735–744 (2016).
101. Xu, L. et al. MYD88 L265P in Waldenstrom macroglobulinemia, immunoglobulin M monoclonal gammopathy, and other B-cell lymphoproliferative disorders using conventional and quantitative allele-specific polymerase chain reaction. *Blood* **121**, 2051–2058 (2013).
102. Kasar, S. et al. Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat. Commun.* **6**, 8866 (2015).
103. Kim, J. et al. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat. Genet* **48**, 600–606 (2016).
104. Boiarsky, R. et al. Single cell characterization of myeloma and its precursor conditions reveals transcriptional signatures of early tumorigenesis. *Nat. Commun.* **13**, 7040 (2022).
105. Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847 (2007).

Acknowledgements

We would like to thank the patients who participated in this study. Anna V. Justis, PhD, a medical writer employed by Dana-Farber Cancer Institute, contributed to this manuscript in part, under the direction of the authors. This research was supported in part by the National Institutes of Health (R35CA263817-01A1 awarded to I.M.G.), and the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation. This research was supported by the International Waldenstrom's Macroglobulinemia Foundation's (IWMF) Robert A. Kyle Award (awarded to R.S.P.). Opinions, interpretations, conclusions, and recommendations are those of the author(s) and are not necessarily endorsed by the IWMF. This research was also supported by a Stand Up To Cancer-Bristol-Meyers Squibb Catalyst Research Grant (Grant Number: SU2C-AACR-CT05-17). Stand

Up To Cancer is a division of the Entertainment Industry Foundation. Research Grants are administered by the American Association for Cancer Research, the Scientific Partner of SU2C. R.S.P. is supported by the Multiple Myeloma Research Foundation Research Fellowship Award, the IWMF's Robert A. Kyle Award, the Dana-Farber Cancer Institute's Center for Early Detection and Interception of Blood Cancers Award, the Claudia Adams-Barr Award for Innovative Basic Cancer Research, the Center of Early Detection and Interception of Blood Cancers at the Dana-Farber Cancer Institute, and the FNIH. Y.K. is supported by a grant from Japan Society for the Promotion of Science; Grant-in-Aid for JSPS Fellow (20J01623), JSPS Overseas Research Fellowships, Mochida Memorial Foundation for Medical and Pharmaceutical Research, and the Elsa U. Pardee Foundation.

Author contributions

R.S.P., Y.K., I.M.G., and G.G. conceived and designed the study; D.H.M., K.T., L.H., N.T., and C.I.L. were in charge of sample collection and processing; R.S.P., Y.K., D.H.M., and N.T. acquired the data; R.S.P., Y.K., D.H.M., T.W., M.P.A., J.T., N.J.H., and E.D.L. analyzed the data; Z.R.H. processed the bulk RNA-seq data; A.K.A., E.B., R.R., and E.L.S. provided guidance in NK cell functionality experiments; R.S.P., N.J.H., and G.G. provided guidance in sequencing data analysis; R.S.P., Y.K., E.K., M.A.D., S.P.T., G.G., and I.M.G. interpreted the data; R.S.P. and Y.K. drafted the manuscript; all authors reviewed, edited and approved the manuscript.

Competing interests

Y.K., D.H.M., T.W., N.T., M.P.A., Z.R.H., A.K.A., J.T., N.J.H., E.D.L., K.T., L.H., E.L.B., and C-I.L. have no conflicts to disclose. R.S.P. is a consultant, equity holder, and co-founder of Predicta Biosciences. M.A.D. has received honoraria from participation in advisory boards and satellite symposia for Amgen, Sanofi, Regeneron, Menarini, Takeda, GlaxoSmithKline, Bristol Myers Squibb, Janssen, BeiGene, Swixx, and Astra Zeneca. R.R. is a co-founder of InnDura Therapeutics and on the scientific advisory board for Glycostem Therapeutics. E.L.S. has consulted for ArsenalBio, Blackstone Life Sciences, Chroma Medicine, Clade Therapeutics, Eureka Therapeutics, ImmuneBridge, Legend Biotech, Overland Pharmaceuticals, Predicta Biosciences and Sana Biotech; is an inventor on licensed patents (US10821135B2, granted; US10633426B2, granted; US10590196B2, granted; US11066475B2, granted; US20200123250A1, granted) that are unrelated to this work and receives royalties from Bristol Myers Squibb and Sanofi; is on the Scientific Advisory Boards of Bristol Myers Squibb, Chimeric Therapeutics and Sanofi; holds equity in Predicta Biosciences and receives research funding from Sanofi. E.K. has received honoraria from Janssen, GlaxoSmithKline, Pfizer, and Sanofi, and research support from Janssen, GlaxoSmithKline, and Pfizer. S.P.T. is a consultant for and receiving research funds from AbbVie/Pharmaceuticals, Janssen, BeiGene, Eli Lilly, and Bristol Myers Squibb. G.G. is a consultant, a current equity holder, and a co-founder for Scorpion Therapeutics and Predicta Biosciences, is receiving research funds from IBM, Pharmacyclics, and is a holder of patents and royalties for SignatureAnalyzer-GPU (US 2021/0358574, published), MSMuTect and MSMutSig (US 11,608,533, issued), MSIDetect (US 2023/0332246, published), and POLYSOLVER (US 11,725,237, issued). None of these patents are related to this work. I.M.G. is a consultant for AbbVie, Adaptive, Amgen, Bristol Myers Squibb, Takeda, Janssen, Novartis, Vor Biopharma (speakers bureau), Sanofi, Pfizer, Menarini Silicon Biosystems, GlaxoSmithKline, the Binding Site, Aptitude Health, Huron Consulting, Oncopeptides, Window Therapeutics, Regeneron, and 10X Genomics; and a consultant, equity holder, board member, and co-founder of Predicta Biosciences. I.M.G.'s spouse, William Savage, MD, PhD, is CMO and equity holder of Disc Medicine. R.S.P., G.G., and I.M.G. are co-inventors on a provisional patent application on the use of gene expression signatures and single-cell immune profiling for the diagnosis and risk stratification of WM.

Ethical approval

One or more of the authors of this paper self-identifies as an under-represented ethnic and/or gender minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community. We support inclusive, diverse, and equitable conduct of research.

Additional information

Supplementary information The online version contains supplementary material available at

<https://doi.org/10.1038/s41467-025-56323-w>.

Correspondence and requests for materials should be addressed to Irene M. Ghobrial.

Peer review information *Nature Communications* thanks Minghao Dang, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

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

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025