ARTICLE
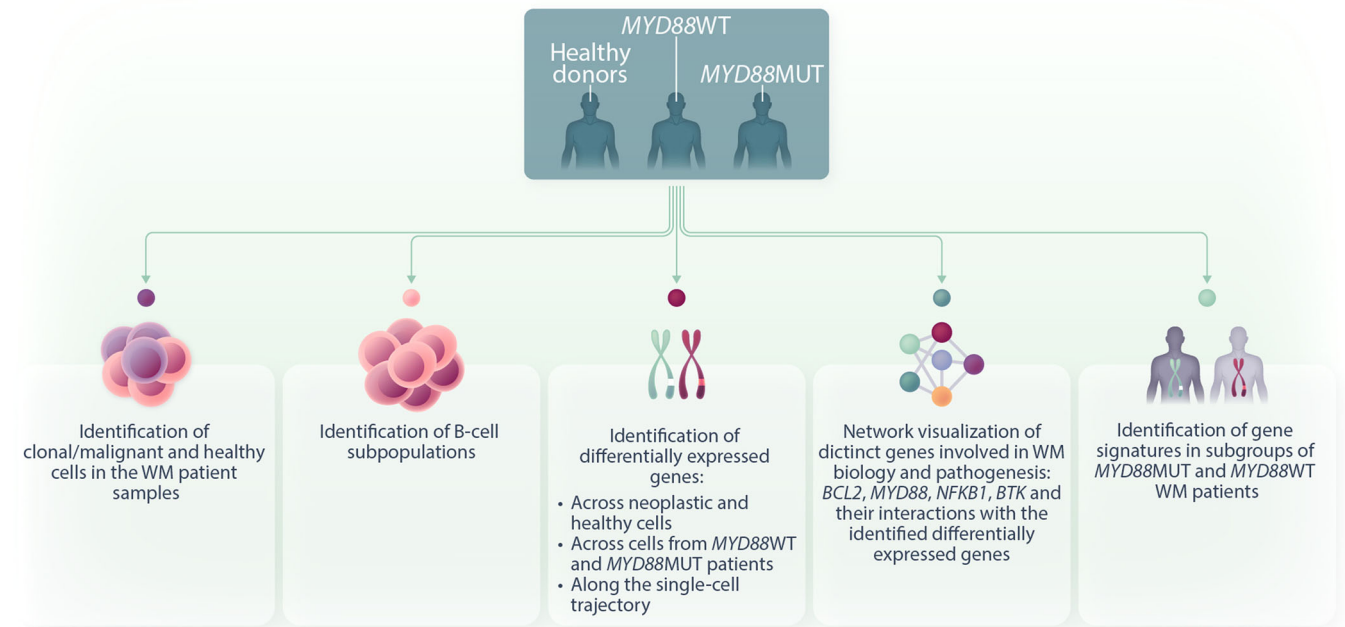
HemaSphere ★EHA EUROPEAN HEMATOLOGY ASSOCIATION

Single-cell analysis of MYD88[L265P] and MYD88[WT] Waldenström macroglobulinemia patients

Tina Bagratuni[1] ⓘ | Foteini Aktypi[1] | Ourania Theologi[1] | Maria Sakkou[2,3,4] |
Kleio Maria Verrou[2,5] | Nefeli Mavrianou-Koutsoukou[1] | Dimitrios Patseas[1] |
Christine Liacos[1] | Stamatia Skourti[1] | Alexandra Papadimou[1] |
Kostantina Taouxi[1] | Foteini Theodorakakou[1] | Georgios Kollias[2,3,4] |
Petros Sfikakis[2,5] | Evangelos Terpos[1] | Meletios A. Dimopoulos[1] |
Efstathios Kastritis[1]

Graphical Abstract

Single-cell RNA-seq of human CD19+ sorted bone marrow B-cells from Waldenström macroglobulinemia (WM) patients and healthy donors

HemaSphere ★EHA EUROPEAN HEMATOLOGY ASSOCIATION

Single-cell analysis of MYD88[L265P] and MYD88[WT] Waldenström macroglobulinemia patients

Tina Bagratuni[1] (ORCID) | Foteini Aktypi[1] | Ourania Theologi[1] | Maria Sakkou[2,3,4] |
Kleio Maria Verrou[2,5] | Nefeli Mavrianou-Koutsoukou[1] | Dimitrios Patseas[1] |
Christine Liacos[1] | Stamatia Skourti[1] | Alexandra Papadimou[1] |
Kostantina Taouxi[1] | Foteini Theodorakakou[1] | Georgios Kollias[2,3,4] |
Petros Sfikakis[2,5] | Evangelos Terpos[1] | Meletios A. Dimopoulos[1] |
Efstathios Kastritis[1]

Correspondence: Tina Bagratuni (tbagratuni@med.uoa.gr)

Abstract

Waldenström macroglobulinemia (WM) is characterized by the expansion of clonal lymphoplasmacytic cells; the MYD88L265P somatic mutation is found in >90% of patients, but malignant B cells may still display intra-clonal heterogeneity. To assess clonal heterogeneity in WM, we generated and performed single-cell RNA sequencing of CD19+ sorted cells from five patients with *MYD88[L265P]* and two patients with *MYD88[WT]* genotype as well as two healthy donors. We identified distinct transcriptional patterns in the clonal subpopulations not only between the two genetically distinct WM subgroups but also among *MYD88[L265P]* patients, which affected the B cell composition in the different subgroups. Comparison of clonal and normal/polyclonal B cells within each patient sample enabled the identification of patient-specific transcriptional changes. We identified gene signatures active in a subset of MYD88[L265P] patients, while other signatures were active in *MYD88[WT]* patients. Finally, gene expression analysis showed common transcriptional features between patients compared to the healthy control but also differentially expressed genes between *MYD88[L265P]* and *MYD88[WT]* patients involved in distinct pathways, including NFκB, BCL2, and BTK. Overall, our data highlight the intra-tumor clonal heterogeneity in WM with potential prognostic and therapeutic implications.

INTRODUCTION

Waldenström's macroglobulinemia (WM) is a low-grade B-cell lymphoproliferative disorder characterized by bone marrow (BM) infiltration by lymphoplasmacytic lymphoma (LPL) cells that secrete IgM immunoglobulin.[1,2] Our understanding of the pathogenesis of WM has improved significantly by the discovery that the vast majority of WM patients harbor the somatic mutation L265P in the *MYD88* gene in their clonal cells.[3–8] However, a small fraction of patients (3%–10%) who lack mutations in *MYD88* may have a different clinical course and disease characteristics, with an increased risk of disease transformation.[9–11] MYD88[L265P] is associated with tonic activation of the BCR pathway and the first FDA-approved anti-WM therapeutic, ibrutinib, which blocks BTK and HCK activity, has shown to be less effective in *MYD88[WT]*, suggesting differences between the two WM variants.[12–14] In addition, WM is a clonal disease but there is heterogeneity in the depth of response to BTK inhibition suggesting that there may be significant heterogeneity within the clone, which has not been extensively explored. A thorough characterization of these genetically distinct

[1]Department of Clinical Therapeutics, School of Medicine, National and Kapodistrian University of Athens, Athens, Greece

[2]School of Medicine, Center of New Biotechnologies & Precision Medicine, National and Kapodistrian University of Athens, Athens, Greece

[3]Department of Physiology, National and Kapodistrian University of Athens Medical School, Athens, Greece

[4]Biomedical Sciences Research Center (BSRC) 'Alexander Fleming', Institute for Bioinnovation, Vari, Greece

[5]Joint Rheumatology Program, National and Kapodistrian University of Athens Medical School, Athens, Greece

subgroups of WM patients could be useful to distinguish the clonal architecture of WM, the molecular mechanisms that underlie resistance to therapy, and the risk of disease transformation as well as to identify potential therapeutic targets.

Recent advances in single-cell RNA sequencing (scRNA-seq) technologies have allowed to sequence and analyze thousands of cells delivering insights into a tumor's cellular heterogeneity and the biological features that distinguish different cell populations. In this study, we generated and analyzed scRNA-seq data from 40,562 single B cells representing seven patients with WM, of which five harbored $MYD88^{L265P}$ ($MYD88^{MUT}$) and two harbored $MYD88^{WT}$ genotype. Our analysis also included two healthy donors (HDs). We explored transcriptional changes within clonal cells, utilized computational methods for separating clonal cells from polyclonal cells, and highlighted gene signatures that are prominent in each subgroup of patients. This revealed insights into the intratumor heterogeneity within and between $MYD88^{WT}$ and $MYD88^{MUT}$ WM patients.

MATERIALS AND METHODS

Bone marrow aspirates were collected from five newly diagnosed $MYD88^{MUT}$ symptomatic WM patients, two newly diagnosed $MYD88^{WT}$ symptomatic WM patients, and two HDs (patient characteristics in Supporting Information S1: Table 1); Sorting of $CD19^+$ cells was performed as previously described[15] (Supporting Information S1: Supplemental Methods). The samples for the single-cell analysis were then prepared following the protocol from 10x Genomics and loaded on the Chromium platform for library preparation according to the manufacturer's instructions. Next-generation sequencing was performed on the 10x Genomics Chromium Illumina NextSeq. 550 platform according to Illumina standard procedures. The Cell Ranger software (version 4.0.0) was used to demultiplex raw sequencing data and align reads to the GRCh38 human reference genome using default parameters, and feature count matrices for a single library were produced.[16] Each single-cell expression matrix was further analyzed in the R program environment (version 4.0.5) with
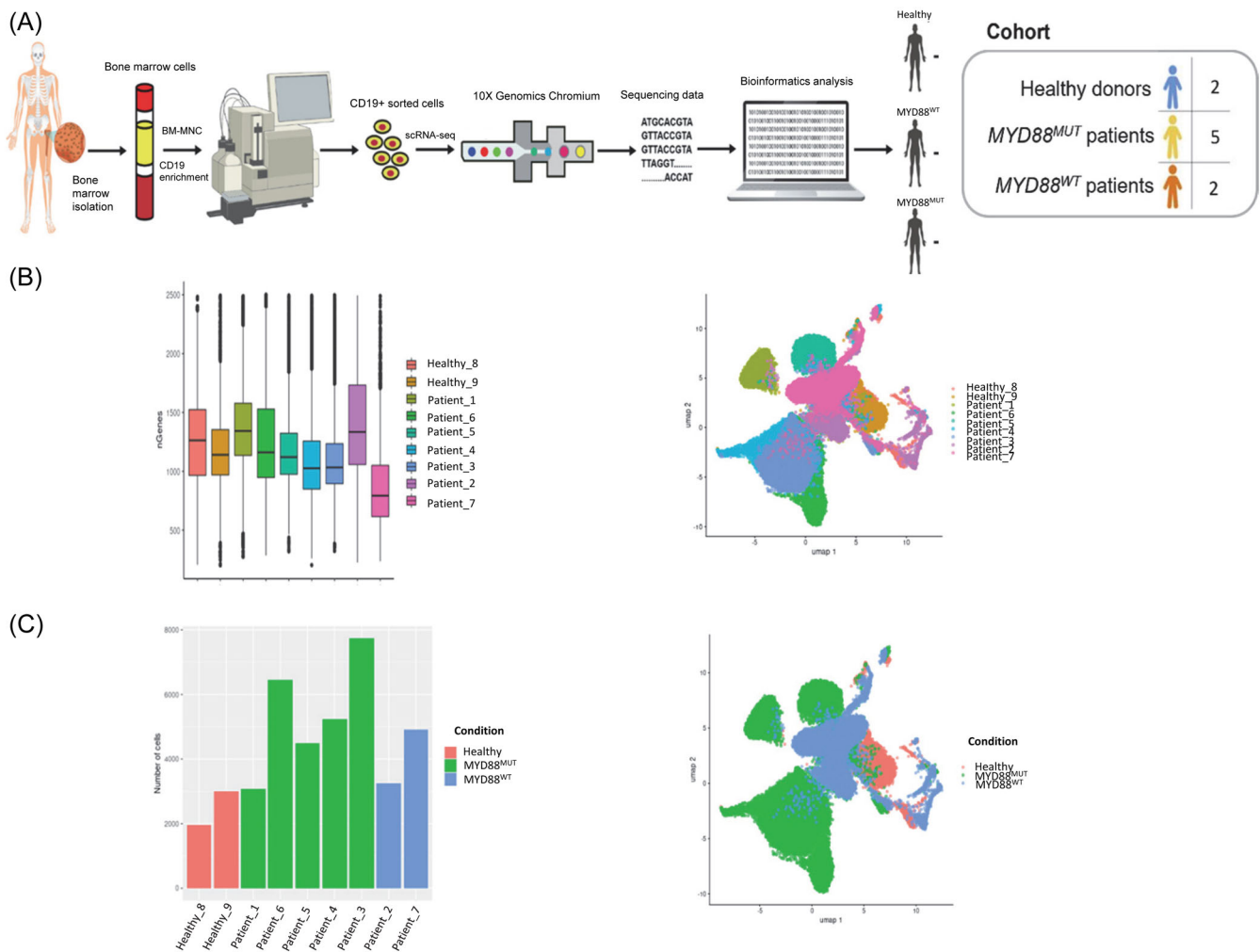

FIGURE 1 Single-cell RNA-seq of human $CD19^+$ bone marrow B cells in Waldenström's macroglobulinemia (WM) patients. (A) Schematic overview of the experimental strategy. (B) Box plot representing the distribution of the number of genes detected per cell in each data set (left) and Uniform Manifold Approximation and Projection (UMAP) visualization of the panoramic integration of the nine single-cell data sets (right) with cells colored according to the donor ID or patient ID. (C) Bar plot of the number of cells per sample (left) and UMAP visualization of the panoramic integration of the nine single-cell data sets (right) with cells colored according to the condition of each individual (healthy, $MYD88^{MUT}$, and $MYD88^{WT}$ WM patients).
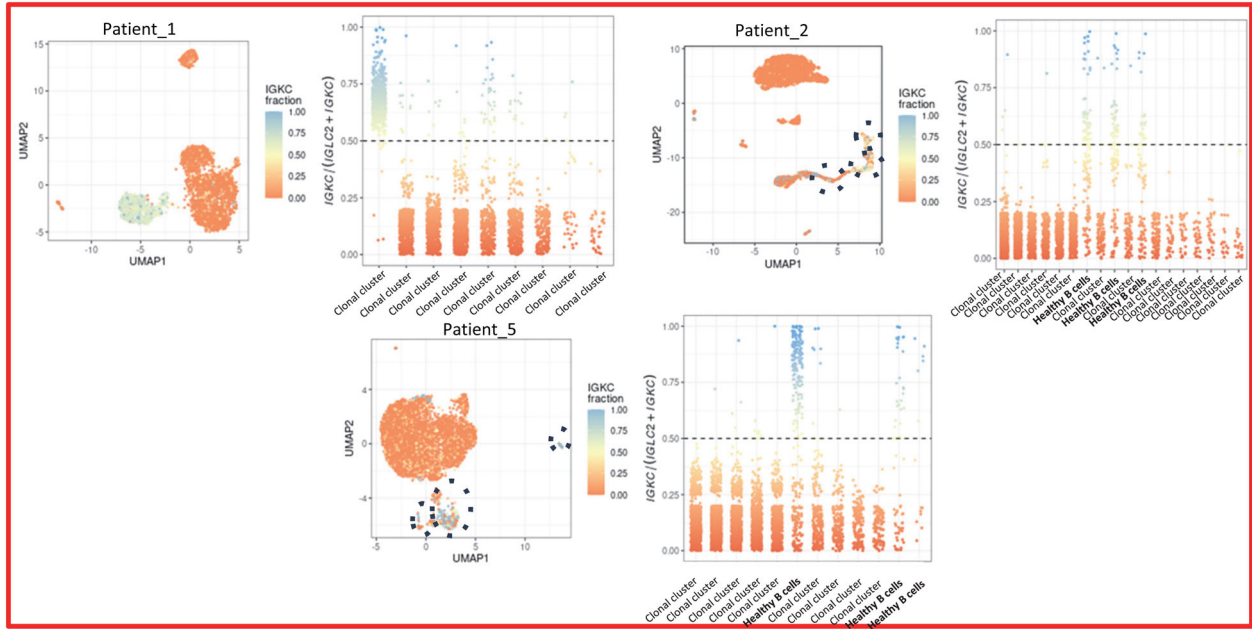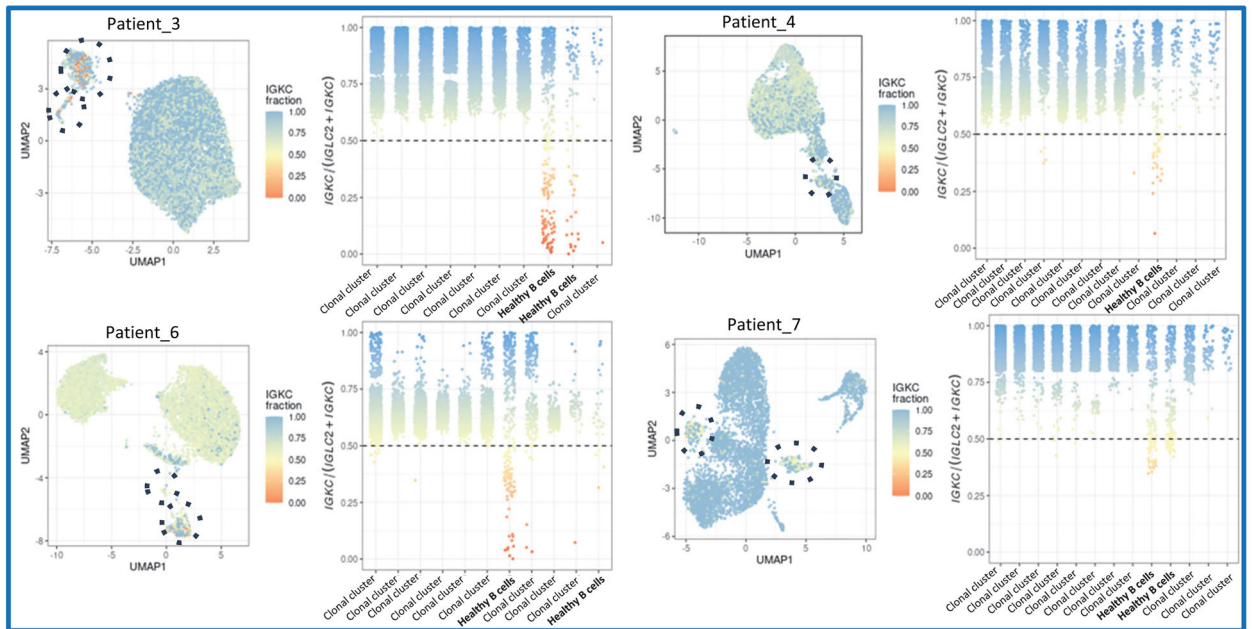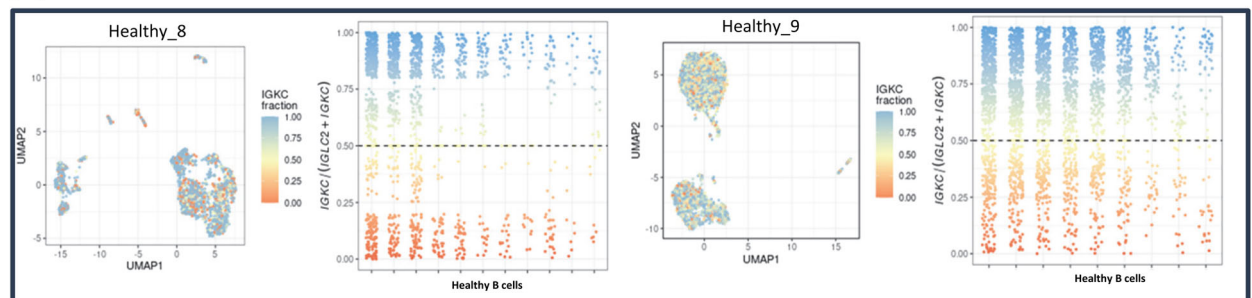
FIGURE 2 (See caption on next page).

FIGURE 2 Identification of malignant and healthy cells in $MYD88^{MUT}$ and $MYD88^{WT}$ Waldenström's macroglobulinemia (WM) patients and healthy donors. For each healthy donor (HD) and WM patient, we present two plots: On the left, a Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction plot coloring each single B cell according to its $IGKC$-fraction: $IGKC/(IGKC^{+}IGLC2)$: kappa-positive B cells ($IGKC$-fraction > 0.5) are colored in blue and lambda-positive B cells ($IGKC$-fraction < 0.5) are colored in red. On the right, a scatter plot displays the $IGKC$-fraction of each cell along the y axis, with cells assigned to donor/patient clusters along the x axis. The malignant clusters contain B cells homogeneously expressing the kappa or lambda light chain in each patient and are annotated as "Clonal clusters." The nonmalignant healthy B cell (hB) clusters contain both kappa-positive and lambda-positive B cells; these variably colored clusters are circled on the UMAP plots (left) and annotated on the scatter plot (right) as "Healthy B cells." The generated plots of (A) the patients expressing $IGLC$, (B) the patients expressing $IGKC$, and (C) the HDs are presented.

the package "Seurat" (version 4.0.2).[17] Detailed descriptions of the methods can be found in the Supporting Information.

RESULTS

Single-cell transcriptional profiles of the two genetically distinct WM patients

We performed droplet-based single-cell RNA sequencing of the nine CD19^{+} sorted BM samples ($MYD88^{MUT}$, n = 5), $MYD88^{WT}$, n = 2 and HDs, n = 2) (Figure 1A). After filtering cells using standard quality controls, we analyzed a total of 40,087 single CD19^{+} B cells (26,976 from $MYD88^{MUT}$ patients, 8155 from $MYD88^{WT}$ patients, and 4956 from the HD), with an average of 4454.11 (range: 1956–7731) cells for each sample. A total of 19,734 different gene transcripts with an average of 116,688 (median: 1099, range: 202–2499) genes for each cell were identified (Figure 1B).

We applied Scanorama to merge nine scRNA-seq data sets: two from healthy donors (n = 1956, n = 3000 cells), five from $MYD88^{MUT}$ WM patients (n = 3080, n = 6455, n = 4478, n = 5232, n = 7731 cells), and two from $MYD88^{WT}$ WM patients (n = 3242, n = 4913 cells). Merging patient data sets using Scanorama's[18] integrative framework and visualizing cells using Uniform Manifold Approximation and Projection (UMAP) plots, we observed that cells were grouped by condition: the majority of the cells originating from HD samples grouped together, the majority of the cells originating from $MYD88^{WT}$ patients also grouped together while the majority of the cells originating from $MYD88^{MUT}$ patients formed different clusters with a smaller amount of overlapping cells with the other data sets, suggesting heterogeneity within the $MYD88^{MUT}$ cohort (Figure 1C).

Light chain restriction distinguishes clonal from polyclonal B cell population

To further distinguish the clonal (malignant) from the polyclonal (nonmalignant) B cell population in each patient, we characterized each B cell as kappa or lambda positive according to its ($IGKC$) fraction $IGLC$-fraction ($IGKC/(IGLC2^{+}IGKC)$), based on the expression of $IGKC$ and $IGLC2$ genes. The reason we selected $IGLC2$ isotype is due to the fact that it was the only informative IGLC isotype among the rest ($IGLC1$, $IGLC3$, $IGLC4$, $IGLC5$, $IGLC6$, and $IGLC7$) (Supporting Information S2: Figure 1A). The $IGLC2$ gene is one of the functional isotypes among the constant lambda genes in the human genome. Its specific expression plays a significant role in immunoglobulin light chain formation.[19,20] We identified each patient's B cell sub-clusters, and we calculated the ratio of kappa positive to lambda positive B cells (κ/λ ratio) in each B cell subcluster, in order to characterize each patient's subclusters as clonal or nonclonal (healthy). According to serum immunofixation, three out of seven patients were of lambda while four were of kappa isotype. This analysis validated the immunofixation results and clonal B-cells from kappa and lambda

patients clustered uniformly highly expressing kappa and lambda light chain, respectively (Figure 2A,B, Supporting Information S2: Figure 1B–H). As expected, cells from HDs were evenly distributed along the IGKC-fraction axis with an average ratio of 53.86% of kappa-positive B cells (B cells expressing the $IGKC$ gene with $IGKC$-fraction > 0.5) expression typically seen in normal B cells (Figure 2C, Supporting Information S2: Figure 1I–K). Of note, in one IgMλ $MYD88^{MUT}$ patient (Patient_1, Figure 2A), a small cluster of B cells expressing the kappa light chain gene ($IGKC$) was also observed. In retrospect, a reevaluation of immunofixation electrophoresis indicated that a faint k-light chain may have been present. Overall, our results show high rates of tumor cell purity (Figure 2D) (mean: 0.941, median: 0.935, range: 0.888–1) with only a small fraction of polyclonal B cells within the patient samples (Figure 2E). Clonality assessment using κ:λ ratios was also concordant with the FACS results (data not shown). Overall, our results show high rates of tumor cell purity (Supporting Information S2: Figure 2A) (mean: 0.941, median: 0.935, range: 0.888–1) with only a small fraction of polyclonal B cells within the patient samples.

We visualized the expression levels of the heavy chain genes across clusters from both patients and HDs using a heatmap. Hierarchical clustering was also conducted row-wise (gene-wise) and column-wise (cluster-wise), utilizing the scaled and averaged gene expression values. The outcome of our clustering revealed a uniform expression of IGH and IGL genes among clusters within each patient (Supporting Information S2: Figure 2B). Furthermore, we have selected the most aberrantly expressed IGH and IGL genes for each patient and produced violin plots, indicating their expression among the healthy and malignant clusters (Supporting Information S2: Figure 2C).

Annotation of malignant and normal cells was further validated by calculating copy number variations (CNV) using inferCNV scores of each cell. InferCNV is used to explore tumor single-cell RNA-Seq data to identify evidence for large-scale chromosomal CNVs, such as gains or deletions of the entire chromosome or large segments of chromosomes. Our analysis showed that inferCNV scores were significantly higher in malignant cells (inferCNV score = 51.89) compared with healthy B cells (for HD samples and patients, inferCNV score = 38.70) (Supporting Information S2: Figure 2C). Unpaired Wilcoxon's test was performed between the inferCNV scores of malignant and healthy clusters. These findings highlight the ability of sc-RNA seq analysis to dissect clonality.

Sc RNA-seq identifies multiple B cell stages within patients and HDs

To identify shared cell populations among patients and HD, characterize potential intra-cluster cell state changes, and identify the B cell composition of WM patients, we performed an integrated analysis of the B and pro/pre-B cells of all samples. Using Harmony integration analysis and applying Louvain clustering (with resolution

FIGURE 3 (See caption on next page).

FIGURE 3 Integration analysis of human CD19$^+$ bone marrow B cells from Waldenström's macroglobulinemia patients and healthy donors and identification of B cell types present in all data sets. Uniform Manifold Approximation and Projection (UMAP) visualizations after integration (Harmony) of the nine scRNA-seq data sets and clustering (Seurat3 Louvain algorithm) on the integrated data set. (A) Each point represents a cell that is colored according to its cluster. (B) Expression heatmap of the top 10 markers (or all markers if less than 10) for each cluster. Purple to yellow indicates low to high levels of gene expression. (C) Gene feature plots show the expression of significant gene expression markers for each cell. (D) Each cluster is classified into one of the broad B cell subpopulations, ranging from pro-B cells to plasmablasts using canonical markers to match the unbiased clustering to known B cell types. (E) Bar plot representing the percentage of each cell type in all samples.

0.4) based on the single-cell RNA expression profile of pro/pre-B-cells and B cells from all nine samples we jointly visualized them in the UMAP plot (Figure 3A). A heatmap graph showing the scaled average expression of the 10 marker genes in each cluster is shown in Figure 3B, providing a visual representation of clusters that share highly expressed markers. We used unsupervised clustering to partition B cells into transcriptionally distinct subsets, which were then annotated by differentially expressed marker genes. Most cells corresponded to the B cell repertoire based on the gene expression of *MS4A1* (CD20) and *CD79A* genes. Specifically, we identified pro B cells (accounting for 1.31% of total cell counts, characterized by the overexpression of *VPREB1* and *MME* genes, pre B cells (accounting for 2.86% of total cell counts) characterized by the expression of *VPREB1*, *MME*, *DNTT*, and *RAG1* genes, a naïve B cell subpopulation characterized by the overexpression of *IGHD* and *IGHM*, which accounted for 15.42% of the total cell counts, a nonswitched (accounting for 55.07% of total cell counts) characterized by the expression of *IGHM* but not *IGHD* and *CD27*, a switched memory B cell subpopulation (accounting 19.31% of total cell counts) characterized by the expression of *CD27* and *JCHAIN* and a plasmablast subpopulation characterized by the overexpression of *JCHAIN* and *XBP1* genes (Figure 3C). Memory B cells (switched and nonswitched) represented the majority of the B cell population with a frequency of 74.38% of total cells. Based on the expression of known marker genes in addition to the top differentially expressed genes for each cluster, we classified our clusters within 10 broad B cell subpopulations, ranging from pro-B cells to plasmablasts (Figure 3D). A detailed bar plot representing the frequency of each cell type in all samples is shown in Figure 3E.

Pseudotime analysis of B cell development trajectory

To understand the relationship between B cell developmental states and changes in gene expression over the time trajectory, we performed single-cell trajectory analysis using Monocle.[21] Trajectory analysis allowed the determination of the B-cell stage composition within the integrated analysis. Pseudotime values were calculated to establish a developmental trajectory (Figure 4A). The analysis showed that two principal stages, distributed in four branches, took place. More specifically, in the first principal stage, which is only seen in *MYD88^{MUT}* patients, naïve B cells originate from pro/pre-B cells, and transit through the nonswitched memory B cells and the mature/switched memory B-cell state into plasmablasts. In the second stage (seen in HDs and *MYD88^{WT}* patients), naïve B cells still originate from pro/pre-B cells, they proceed to nonswitched memory B cells, however, they do not proceed to switched memory B cells and plasmablasts (only a few cells appear in these stages) (Figure 4B). Once the cells were placed in order according to their pseudotime values, we searched for genes whose expression changes as the cells move along the pseudotime and we grouped together genes that have similar trends. We constructed a heatmap to explore dynamic expression changes of genes associated with cellular transitions (Figure 4E). The group of genes in group 1 is mainly differentially expressed at the early stages of the pseudotime with a gradual decrease as we move along to the next stages and is mainly involved in STAT6, JUNB, and MYC signaling pathways. The group of genes differentially expressed in the middle of pseudotime is mainly involved in IL-2 signaling, TCF4, and MYC signaling pathways. Almost all genes differentially expressed at later stages of the pseudotime (mostly seen in *MYD88^{MUT}* patients) primarily driven by *RPL* genes are involved in the MYC pathway. Finally, at the last stage of the pseudotime, which is mainly seen in *MYD88^{MUT}* patients, the differentially expressed genes are mainly involved in KRAS, TNFα, and p53 pathways. The top 25 Gene Ontology (GO) biology processes are shown in Figure 4F that demonstrate significant enrichment in protein localization to endoplasmic reticulum, protein targeting, and RNA catabolic process.

B cell composition in *MYD88^{WT}* and *MYD88^{MUT}* patients

We next aimed to determine the transcriptional differences seen between the two genetically distinct WM patient groups. To this end, we split the samples into condition-based cohorts (HD, *MYD88^{MUT}*, *MYD88^{WT}*) within the integrated data set to identify distinct cell populations among patients and HD and to characterize potential intracluster cell state changes (Figure 5A). Our results show that the B cell composition of the *MYD88^{MUT}* cohort of patients comprises six B cell subtypes, as described above, while the B cell composition of the *MYD88^{WT}* cohort is mainly restricted to four B cell subtypes. *MYD88^{WT}* cohort seems to resemble the HD cohort, where clusters comprising the switched memory B cells are restricted and the naïve B cell composition is overexpressed compared with other cell types. To further demonstrate the observable differences seen between the two genetically distinct WM groups, we split the samples into individual patients and HDs within the integrated data set (Figure 5B, Supporting Information S2: Figure 3). Our results show that the B cell composition of the HDs and the *MYD88^{WT}* patients (patients 2 and 7) share almost the same transcriptional profile where both HDs mostly express naïve B cell population accompanied with a moderate expression of pre-B cells and a cluster of nonswitched memory B cells while both *MYD88^{WT}* patients mostly express pre/pro, naïve and nonswitched memory B cells with a median expression of plasmablasts. In contrast, *MYD88^{MUT}* patients seem to have a more heterogeneous transcriptional profile, which can be divided into two subgroups. Three out the five *MYD88^{MUT}* patients (patients 3, 4, and 6) have an enriched population of switched memory B cells and plasmablasts in addition to a cluster of nonswitched memory B cells but with a minimal population of naïve B cells while two out of the five *MYD88^{MUT}* patients (patients 1 and 5) have no population of switched memory B cells and plasmablasts but are more enriched in naïve and part of nonswitched memory B cells. Pro/pre-B cells in all five *MYD88^{MUT}* patients are almost absent. Switched memory B cells demonstrated significant expression for genes, such as *NEB*, *PCDH9*, *MAN1A1*, *FGL2 HCK*, *CCND2*, and *MARCKS* (Figure 5C).
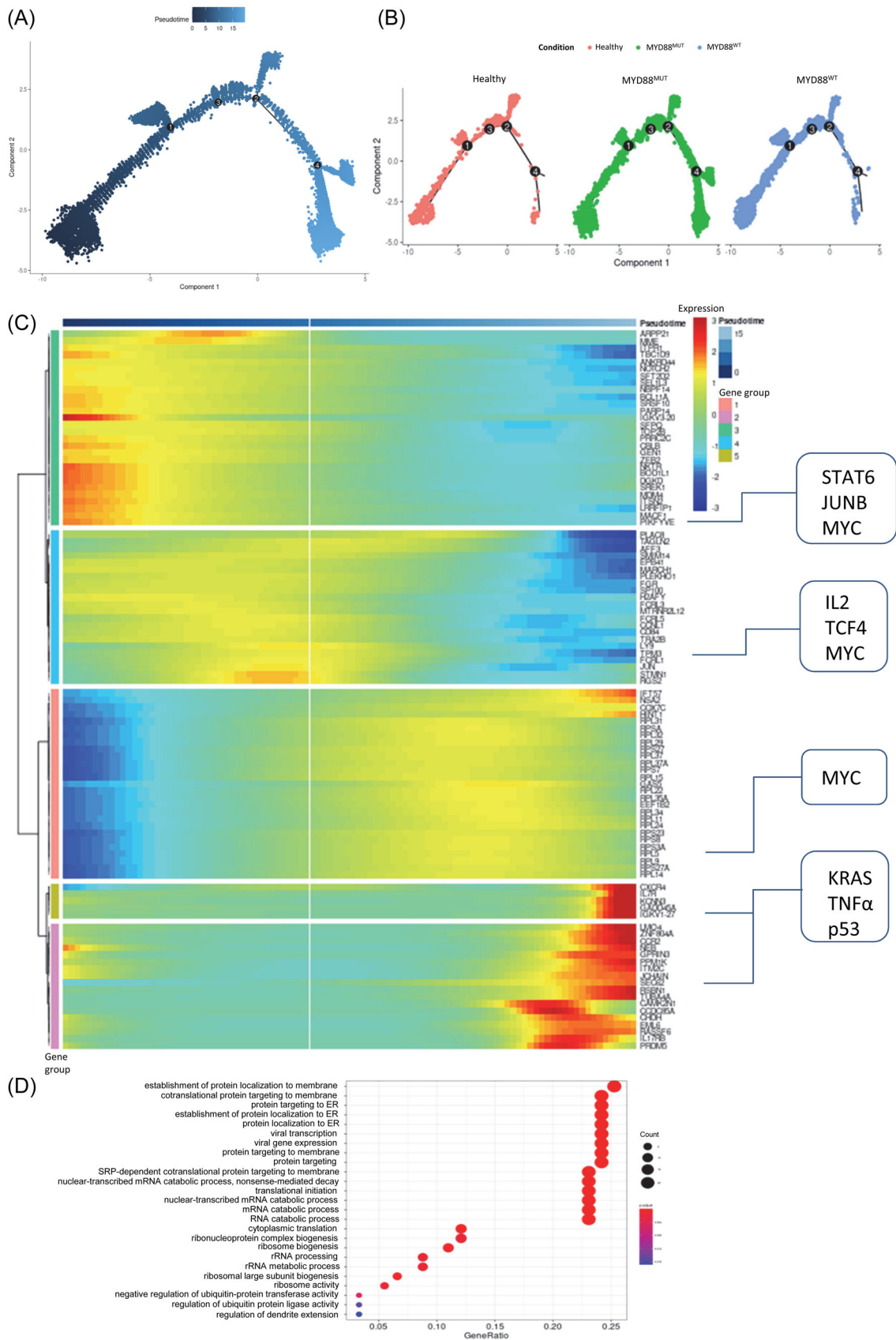
FIGURE 4 (See caption on next page).

FIGURE 5 B cell type composition of the $MYD88^{MUT}$ and the $MYD88^{WT}$ Waldenström's macroglobulinemia patients. Uniform Manifold Approximation and Projection (UMAP) visualization of the integrated data set (A) according to the condition (healthy donor [HD], $MYD88^{MUT}$, or $MYD88^{WT}$) and (B) in individual patients and HDs. Each point represents a cell colored according to its B cell type. Violin plots display the expression of representative B cell markers in each cell type.

FIGURE 6   (See caption on next page).

FIGURE 6 Differential gene expression across neoplastic and healthy cells (pseudobulk approach). (A) MA plot representing log2 fold-change (*y* axis) versus log2 mean expression (*x* axis) between healthy and neoplastic cells. Each point represents a gene; the genes with log2 fold-change > 0.25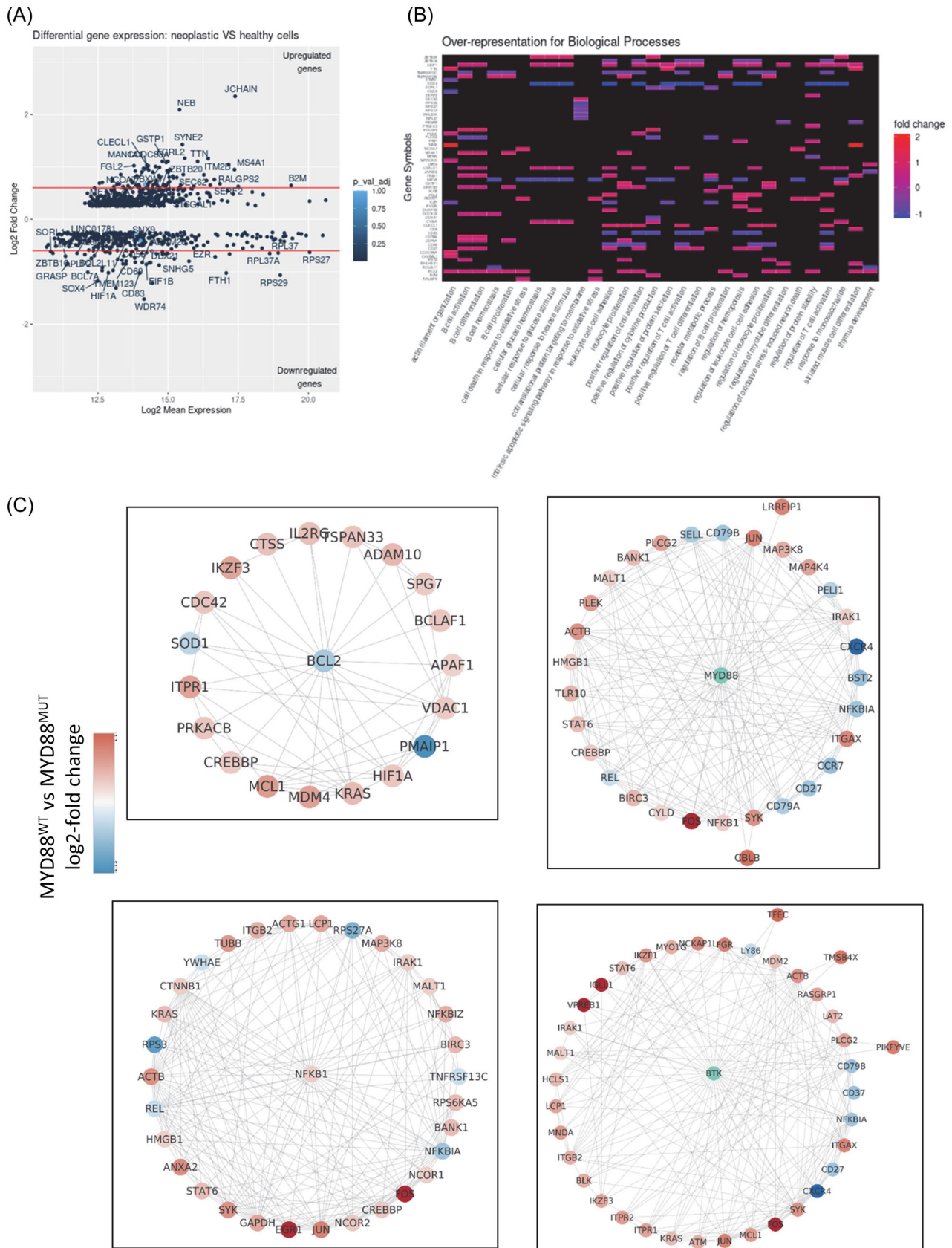 or log2 fold-change < −0.25 have been plotted. Genes with log2 fold-change within the range −0.25 to 0.25 are not shown. Red lines show differential expression thresholds for upregulated genes (log2FoldChange > 0.6) and downregulated genes (log2FoldChange < −0.6). *p* Value threshold: *p* value < 0.05. (B) Heatplot visualization of the Functional Enrichment analysis of the differentially expressed genes between healthy and neoplastic cells. Enrichment Gene Ontology categories after false discovery rate control are shown on the *x* axis. Genes involved in these significant categories are shown on the *y* axis. (C) Network visualizations of distinct genes involved in WM biology and pathogenesis (*BCL2*, *MYD88*, *NFKB1*, and *BTK*) and their interactions with other differentially expressed genes across cells from *MYD88^WT^* and *MYD88^MUT^* patients. Gray lines indicate protein–protein interactions of gene products using experimentally validated interactions from the STRING database.

Transcriptional differences between clonal and normal B cells across patients and HDs

To investigate the differences between clonal and normal B cells, we performed differential expression (DE) analysis to detect differences in gene expression between the clonal and the normal B cells previously identified. We compared the clonal to normal/polyclonal subpopulation using Seurat's function FindMarkers and found 132 differentially expressed genes (DEGs) excluding genes that are related to Ig chains with log2FoldChange < −0.6 (downregulated genes), log2FoldChange > 0.6 (upregulated genes), and *p* value < 0.05. Our analysis showed upregulation of genes typically expressed in WM B cells, such as *MS4A1*, *CD79A*, *CD79B*, and *BCL2*. In addition to these genes, our analysis identified genes that have been less well characterized in WM. The top five upregulated DEGs by log2FC value included *JCHAIN*, *SYNE2*, *ITM2B*, *NEB*, and *TTN* (Figure 6A). The *JCHAIN* gene is involved in creating the binding site for pIgR/SC in the Ig polymers by interacting directly with the receptor protein.[22,23] Therefore, *JCHAIN* plays a key role in secretory immunity.[24] *SYNE2* is involved in cyclin-dependent kinase inhibition and potentially associated with B cell signaling or B cell malignancy. *ITM2B* is a target of BCL6 repression in lymphoma, which is crucial for germinal center B cell development.[25] The top downregulated genes are *SOX4*, *NEIL1*, *BCL7A*, *CD83*, and *HIF1A* (Figure 6A). *SOX4* is important in maintaining the survival of pro-B cells[26] while the *NEIL1* gene participates in the DNA repair pathway and deficiency of the *NEIL1* gene has demonstrated reduced germinal center B cell activity.[27] BCL7A protein is highly expressed in the nuclei of GC B lymphocytes, while mutations in the *BCL7A* gene are recurrently found in DLBCL.[28–30] *CD83* gene is located in chromosome 6q and is commonly affected in WM[10] while *HIF1A* expression is high in pro-B and pre-B cells and decreases at the naive B cell stage.[31] Heatmap analysis showed that the main biological processes of the DE genes include B cell activation, B cell differentiation, and regulation of hemopoiesis (Figure 6B). GO comparative analysis showed that distinct pathways, including NFκB, BCL2, BTK, and MYD88 pathways display differentially regulated genes in the *MYD88^MUT^* compared with the *MYD88^WT^* clones, which could provide better insights into specific genes within these pathways that may be involved in disease transformation and response to therapy (Figure 6C).

*MYD88^MUT^* and *MYD88^WT^* clonal B cells are transcriptionally distinct from normal B cells within patients

In order to evaluate intratumor heterogeneity, we performed DE analysis of malignant and normal B cell populations within our patient cohort; that is, for each patient we compared the malignant B cells to their own healthy B cells. This allowed us to individually characterize the unique transcriptional landscape of each patient, which may not

be shared across all patients, giving the ability to reduce potential confounding effects that would occur when comparing neoplastic cells to normal cells of HDs. Six patients (four *MYD88^MUT^* and two *MYD88^WT^*) had both malignant and normal B cell populations; one *MYD88^MUT^* patient had only clonal B cells. From the above six patients, five had significant DEGs between populations with log2FoldChange < −0.6 (downregulated genes) or log2FoldChange > 0.6 (upregulated genes) and *p* value < 0.05) (Figure 7A). Overall, our analysis identified 899 DEGs, of which only 92 genes (10.2%) were found in the bulk malignant versus normal DE analysis described above (Figure 7B). We found the *PCDH9* gene to be exclusively differentially expressed in individual patients but also genes that are mutually affected across a number of patients, such as the upregulation of *JCHAIN*, the upregulation of *NEB*, and the downregulation of *SOX4* and *BCL7A*. *PCDH9* gene has been previously reported as being downregulated in nonnodal mantle cell lymphoma and glioblastoma as a result of gene copy number alterations.[32,33] Furthermore, 16 common genes were identified to be commonly expressed in *MYD88^WT^* patients, including genes, such as *CD84*, *FGR*, *PTMA*, *H3F3A*, *FCRL3*, *EGR1*, *FCRER2*, and *PLD4*, which were not found in *MYD88^MUT^* patients (Figure 7C). Notably, in comparison to the general malignant versus normal DEG analysis, we identified DEGs that were commonly upregulated or downregulated in *MYD88^MUT^* patients and were not affected in *MYD88^WT^* patients and vice versa. Hence, we identified the upregulation of *JCHAIN* in two out of four *MYD88^MUT^* patients which was downregulated in both *MYD88^WT^* patients. Furthermore, we identified that *ITM2B*, *SYNE2*, *MAN1A1*, *FGL2*, and *GLCC1* were upregulated (log2FC > 0.6) in three out of four *MYD88^MUT^* patients and in none of the *MYD88^WT^* patients. *MAN1A1* gene is located in chromosome 6q and has been found to be commonly affected in WM patients due to the 6q clonal loss. Regarding the *FGL2* gene, it has been shown to be highly overexpressed in a subset of WM cases, suggesting that WM cells may be responsive to cytokine signaling through T-cell receptors.[34,35] From this analysis, we also identified the downregulation of the *NEIL1* gene in *MYD88^WT^* patients but not in *MYD88^MUT^* patients. Furthermore, the *ARID5B* gene was downregulated in three out of four *MYD88^MUT^* patients and in none of the *MYD88^WT^* patients. Finally, the *CXCR4* gene was upregulated in two out of four *MYD88^MUT^* patients and downregulated in one out of two *MYD88^WT^* patients although all patients harbored *CXCR4^WT^* genotype.

Furthermore, using Bayesian Non-Negative Matrix Factorization (BNMF), we extracted gene expression signatures that were active in our cohort. The BNMF approach helped us to identify groups of genes with shared activity in single cells and examine how signature activity varies between patient samples and between subpopulations of cell clusters in our data set. We highlighted three interesting gene signatures across all B cells in our cohort, each of which represents a pattern of gene expression recurrently occurring across cells in our data set (Figure 7D, Table 1). In addition, we observe a higher expression of the *BCL2L11* gene in the cell populations of HDs while the
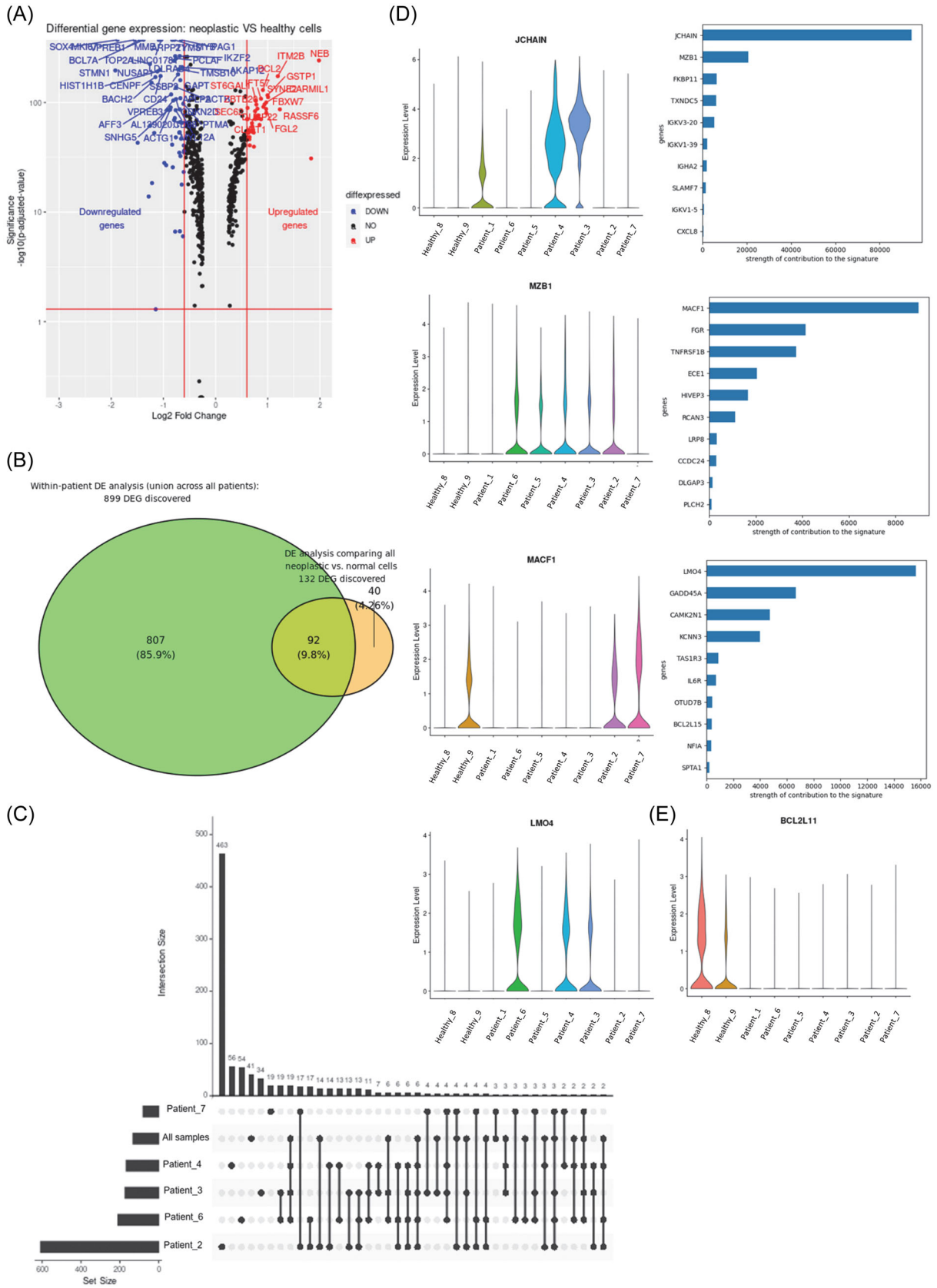
(A)

(B)

(C)

(D)

(E)

FIGURE 7 (See caption on next page).

FIGURE 7 Quantification of DEGs within each patient's neoplastic versus healthy cells and in the general neoplastic versus normal cells. (A) Volcano plot to represent log2 fold-change (x axis) versus statistical significance (−log10(p-adjusted value)—y axis) between healthy and neoplastic cells from only the WM patients. Each point represents a gene; the genes with log2 fold-change > 0.25 or log2 fold-change < −0.25 have been plotted. Genes with log2 fold-change within the range −0.25 to 0.25 are not shown. Red lines show our differential expression thresholds for upregulated genes (log2FoldChange > 0.6) and downregulated genes (log2FoldChange < −0.6) and the significance threshold: −log10(p-adjusted-value) < 0.05. (B) Venn diagram representing the overlap of DEGs using the within-patient DE approach with the DEGs found using the pseudobulk DE approach comparing all neoplastic to healthy cells (derived from the patients only). (C) UpSet plot displays the intersections between the DE gene sets in each patient and in all cells. The rows correspond to the sets and the columns correspond to the intersections. (D) Top 10 genes for three selected gene signatures. Each gene's strength of contribution is plotted on the x axis. Signatures with top contributions from *JCHAIN, MZB1, MACF1, LMO4*. Violin plots indicate in which patients each gene is most active. (E) Violin plot indicating *BCL2L11* gene activity in healthy donors and patients.

TABLE 1 Highlighting signatures associated with disease biology.

Signature top genes	Description	Patients
JCHAIN, MZB1, FKBP11, TXNDC5	*KLF12* and *FOXP1* function	All
MACF1, FGR, TNFRSF1B, ECE1, HIVEP3	*RUNX1* function	MYD88^{WT}
LMO4, GADD45A, CAMK2N1, KCNN3, TAS1RE, IL6R	Normal hematopoiesis	MYD88^{MUT}: Patient_3, Patient_4, Patient_6.

distribution probability of patient cells expressing *BCL2L11* is extremely low (Figure 7E).

DISCUSSION

In this study, we have used sc-RNA sequencing technology to investigate the intratumor heterogeneity of WM B cells in MYD88^{MUT} and MYD88^{WT} patients. To the best of our knowledge, few studies have performed single-cell transcriptome analysis in WM patients; however, this is the first study to focus on the transcriptome of MYD88^{WT} and MYD88^{MUT} patients at the single-cell level.[36,37] Most importantly the characterization of MYD88^{WT} WM, an infrequent subtype of WM, is particularly interesting as it has been shown to be associated with increased risk for disease transformation and relative resistance to ibrutinib. Hence understanding the molecularly driven underlying mechanisms of these distinct genetic subtypes of WM patients is critical for developing informed criteria for patients who would benefit from alternative therapies and frequent monitoring. We analyzed the transcriptomes of individual cells from seven WM and two HD, sampling on average 4500 cells per sample. This allowed significant observations regarding clonal heterogeneity. Notably, among MYD88^{MUT} patients the transcriptome profile was more heterogeneous compared with that of patients with MYD88^{WT} genotype as visualized by the clustering proximity where the majority of the cells originating from the MYD88^{MUT} samples formed three relatively remote clusters while cells from the MYD88^{WT} patients almost clustered in one group of cells.

Precise labeling of normal and neoplastic B cells in each WM patient is a significant challenge since conventional sorting of B cells from the BM will almost always include some normal B cells with the same Ig light chain as the tumor. However, using the sc-RNA seq-based tumor B cell classification with the kappa/lambda light chain fraction ratio restriction, we were able to directly compare the transcriptomes of a purely malignant to a normal B cell population from the same sample. This overcomes to a significant degree the limitations of studies performing bulk analysis, which are likely influenced by the contamination from normal B cells. With sc-RNA seq the distinction of normal versus neoplastic B cells within individual patients was applicable. Through the distinction of clonal from polyclonal cells, we were able to identify genes that were differentially

expressed, which were not discovered by the "bulk" clonal versus normal B cell DE analysis, such as *FGR, CD84,* and *FCER2,* which are involved in the BTK pathway.

To characterize potential intra-cluster cell state changes, using the integrated analysis in the entire cohort of patients and HDs, we identified ten clusters representing six transcriptionally distinct B cell stages. We observed that the B cell composition in the MYD88^{MUT} patients varies: a group of patients is characterized by the expression of all six B cell stages while another group mainly expresses four B cell stages, almost missing the switched memory B cell and plasmablast population. From the clinical point of view, our MYD88^{MUT} patients (n = 2) with restricted switched memory B cell and plasmablast population seem to have a more inferior response to ibrutinib therapy (in terms of IgM response) compared with the MYD88^{MUT} patients expressing all six B cell stages (n = 3). In contrast, the B cell composition in MYD88^{WT} in the two patients was almost homogeneous and transcriptionally comparable to the subgroup of MYD88^{MUT} patients with restricted switched memory B cell and plasmablast population, mainly through the expression of *CD27* and *JCHAIN* genes. *CD27*, one of the top genes of the switched memory B cell signature, has been previously discussed in WM literature with variable expression in WM cells.[38] Studies have shown that WM malignant B cells are scattered throughout CD20$^+$CD27$^+$ and CD20$^+$CD27$^-$ B-cell compartments, with the most residing in the CD20$^+$CD27$^-$ subset.[39] Another study showed that somatically hypermutated WM tumor cells can lack CD27 expression, raising suggestions of unusual memory B-cell origins.[40] On the other hand, *JCHAIN* is expressed mainly in plasma cells regardless of immunoglobulin isotype;[41–43] however, this molecule has not been studied extensively, partly due to technical limitations. The role of *JCHAIN* is to join the five monomeric subunits of secreted pentamer IgM in humans. Hexameric IgM was discovered in patients with different disorders like WM[44,45] and unlike the pentameric IgM, hexamer is characterized by the absence of *JCHAIN*.[46] Hence the DE of *JCHAIN* and *CD27* genes between tumor cells in our cohort of patients might provide some clinical insights into these patients; however, a larger cohort with sc-data needs to be tested to verify these results.

CXCR4 is a key gene in WM biology and mutations are present in almost 40% of patients with WM[10,47,48] with important clinical implications, including ibrutinib resistance.[49] Our sc-data show that *CXCR4* expression varies not only among MYD88^{WT} and MYD88^{MUT}

patients but also among $MYD88^{MUT}$ patients. $CXCR4$ expression is mainly detected in the switched memory B cell subpopulation with minimal expression in the naive and pro/pre-B cell subpopulations. Hence two out of the three $MYD88^{MUT}$ patients with distinct expression of switched memory B cell subpopulation were coupled with high $CXCR4$ expression while in one out of the two $MYD88^{WT}$ patients, $CXCR4$ expression was downregulated. Given the fact that patients harboring $MYD88^{L265P}$ mutation express significantly higher levels of $CXCR4$ compared with $MYD88^{WT}$ patients,[50,51] this might provide us some biological insights on $CXCR4$ expression-dependent ibrutinib response since all patients included in this analysis were wild type for $CXCR4$ mutations.

In terms of B cell development trajectory, our analysis showed that the harmony integrated and the monocle data set platforms have a high degree of overlap in terms of the B cell composition and expected timing of the developmental stages, but more importantly, that specific pathways are activated at earlier stages of B cell differentiation, such as MYC, $JUNB$, and $STAT6$, whereas pathways involving $TP53$, $KRAS$, and $TNF\alpha$ are activated at late stage differentiation of B cell development.

In summary, our single-cell data in WM indicates unique gene expression signatures differentiating the two major genetically distinct types of WM patients. It also reveals heterogeneity among the intra-tumor cell populations, which could explain differential responses to certain drugs (such as BTK inhibitors or anti-CD20 MAbs). Despite the small number of patients, we believe that our study is of high clinical relevance. However, further studies are required to decipher the clinical importance of the observed heterogeneity and to provide further mechanistic insights into the differences in the biology and clinical presentation of patients with $MYD88^{WT}$ versus those harboring $MYD88^{MUT}$.

AUTHOR CONTRIBUTIONS

Designed the research and wrote the paper: Tina Bagratuni. *Analyzed data*: Foteini Aktypi, Ourania Theologi, and Kleio Maria Verrou. *Performed research*: Maria Sakkou, Nefeli Mavrianou-Koutsoukou, Dimitrios Patseas, Christine Liacos, Stamatia Skourti, Alexandra Papadimou, and Kostantina Taouxi. *Analyzed patient data*: Foteini Theodorakakou. *Wrote the paper*: Georgios Kollias, Petros Sfikakis, Evangelos Terpos, Meletios A. Dimopoulos, and Efstathios Kastritis.

CONFLICT OF INTEREST STATEMENT

ET is a HemaSphere editor. The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in GSE235723 at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE235723.

FUNDING

ORCID

Tina Bagratuni http://orcid.org/0000-0001-8695-0493

SUPPORTING INFORMATION

Additional supporting information can be found in the online version of this article.

REFERENCES

1. Owen RG, Treon SP, Al-Katib A, et al. Clinicopathological definition of Waldenstrom's macroglobulinemia: consensus panel recommendations from the Second International Workshop on Waldenstrom's Macroglobulinemia. *Semin Oncol*. 2003;30(2):110-115.

2. Campo E, Swerdlow SH, Harris NL, Pileri S, Stein H, Jaffe ES. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood*. 2011;117(19): 5019-5032.

3. Treon SP, Xu L, Yang G, et al. MYD88 L265P somatic mutation in Waldenström's macroglobulinemia. *N Engl J Med*. 2012;367(9): 826-833.

4. Varettoni M, Arcaini L, Zibellini S, et al. Prevalence and clinical significance of the MYD88 (L265P) somatic mutation in Waldenström's macroglobulinemia and related lymphoid neoplasms. *Blood*. 2013; 121(13):2522-2528.

5. Jiménez C, Sebastián E, Chillón MC, et al. MYD88 L265P is a marker highly characteristic of, but not restricted to, Waldenström's macroglobulinemia. *Leukemia*. 2013;27(8):1722-1728.

6. Xu L, Hunter ZR, Yang G, et al. MYD88 L265P in Waldenström macroglobulinemia, immunoglobulin M monoclonal gammopathy, and other B-cell lymphoproliferative disorders using conventional and quantitative allele-specific polymerase chain reaction. *Blood*. 2013;121(11):2051-2058.

7. Landgren O, Staudt L. MYD88 L265P somatic mutation in IgM MGUS. *N Engl J Med*. 2012;367(23):2255-2257; author reply 2256-2257.

8. Poulain S, Roumier C, Decambron A, et al. MYD88 L265P mutation in Waldenstrom macroglobulinemia. *Blood*. 2013;121(22): 4504-4511.

9. Hunter ZR, Xu L, Tsakmaklis N, et al. Insights into the genomic landscape of MYD88 wild-type Waldenström macroglobulinemia. *Blood Adv*. 2018;2(21):2937-2946.

10. Hunter ZR, Xu L, Yang G, et al. The genomic landscape of Waldenström macroglobulinemia is characterized by highly recurring MYD88 and WHIM-like CXCR4 mutations, and small somatic deletions associated with B-cell lymphomagenesis. *Blood*. 2014;123(11): 1637-1646.

11. Abeykoon JP, Paludo J, King RL, et al. MYD88 mutation status does not impact overall survival in Waldenström macroglobulinemia. *Am J Hematol*. 2018;93(2):187-194.

12. Ansell SM, Kyle RA, Reeder CB, et al. Diagnosis and management of Waldenström macroglobulinemia: Mayo stratification of macroglobulinemia and risk-adapted therapy (mSMART) guidelines. *Mayo Clin Proc*. 2010;85(9):824-833.

13. Treon SP, Tripsas CK, Meid K, et al. Ibrutinib in previously treated Waldenström's macroglobulinemia. *N Engl J Med*. 2015;372(15): 1430-1440.

14. Paulus A, Ailawadhi S, Chanan-Khan A. Novel therapeutic targets in Waldenstrom macroglobulinemia. *Best Pract Res Clin Haematol*. 2016;29(2):216-228.

15. Bagratuni T, Ntanasis-Stathopoulos I, Gavriatopoulou M, et al. Detection of MYD88 and CXCR4 mutations in cell-free DNA of patients with IgM monoclonal gammopathies. *Leukemia*. 2018;32(12): 2617-2625.

16. Zheng GX, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*. 2017;8:14049.

17. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184(13):3573-3587.e29.

18. Berger F, Traverse-Glehen A, Felman P, et al. Clinicopathologic features of Waldenström's macroglobulinemia and marginal zone

lymphoma: are they distinct or the same entity? *Clin Lymphoma*. 2005;5(4):220-224.

19. Abe Y, Sakata-Yanagimoto M, Fujisawa M, et al. A single-cell atlas of non-haematopoietic cells in human lymph nodes and lymphoma reveals a landscape of stromal remodelling. *Nat Cell Biol*. 2022;24(4): 565-578.

20. Roider T, Seufert J, Uvarovskii A, et al. Dissecting intratumour heterogeneity of nodal B-cell lymphomas at the transcriptional, genetic and drug-response levels. *Nat Cell Biol*. 2020;22(7):896-906.

21. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*. 2017;14(10):979-982.

22. Brandtzaeg P, Prydz H. Direct evidence for an integrated function of J chain and secretory component in epithelial transport of immunoglobulins. *Nature*. 1984;311(5981):71-73.

23. Brandtzaeg P. Complex formation between secretory component and human immunoglobulins related to their content of J chain. *Scand J Immunol*. 1976;5(4):411-419.

24. Johansen FE, Braathen R, Brandtzaeg P. Role of J chain in secretory immunoglobulin formation. *Scand J Immunol*. 2000;52(3):240-248.

25. Baron BW, Baron RM, Baron JM. The ITM2B (BRI2) gene is a target of BCL6 repression: implications for lymphomas and neurodegenerative diseases. *Biochim Biophys Acta*. 2015;1852(5):742-748.

26. Schilham MW, Oosterwegel MA, Moerer P, et al. Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. *Nature*. 1996;380(6576):711-714.

27. Mori H, Ouchida R, Hijikata A, et al. Deficiency of the oxidative damage-specific DNA glycosylase NEIL1 leads to reduced germinal center B cell expansion. *DNA Repair*. 2009;8(11):1328-1332.

28. Ramos-Medina R, Montes-Moreno S, Maestre L, et al. BCL7A protein expression in normal and malignant lymphoid tissues. *Br J Haematol*. 2013;160(1):106-109.

29. Reddy A, Zhang J, Davis NS, et al. Genetic and functional drivers of diffuse large B cell lymphoma. *Cell*. 2017;171(2):481-494.

30. Schmitz R, Wright GW, Huang DW, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N Engl J Med*. 2018;378(15): 1396-1407.

31. Burrows N, Bashford-Rogers RJM, Bhute VJ, et al. Dynamic regulation of hypoxia-inducible factor-1α activity is essential for normal B cell development. *Nat Immunol*. 2020;21(11):1408-1420.

32. Wang C, Yu G, Liu J, et al. Downregulation of PCDH9 predicts prognosis for patients with glioma. *J Clin Neurosci*. 2012;19(4): 541-545.

33. Tayrac M, Etcheverry A, Aubry M, et al. Integrative genome-wide analysis reveals a robust genomic glioblastoma signature associated with copy number driving changes in gene expression. *Genes Chromosomes Cancer*. 2009;48(1):55-68.

34. Xu H, Yao F. Microarray-based gene expression analysis identifies potential diagnostic and prognostic biomarkers for Waldenström macroglobulinemia. *Acta Haematol*. 2018;140(2):87-96.

35. Barillé-Nion S, Barlogie B, Bataille R, et al. Advances in biology and therapy of multiple myeloma. *Hematol Am Soc Hematol Educ Program*. 2003;248-278.

36. Sun H, Fang T, Wang T, et al. Single-cell profiles reveal tumor cell heterogeneity and immunosuppressive microenvironment in Waldenström macroglobulinemia. *J Transl Med*. 2022;20(1):576.

37. Qiu Y, Wang X, Yao Y, et al. Single-cell transcriptome analysis reveals stem cell-like subsets in the progression of Waldenström's macroglobulinemia. *Exp Hematol Oncol*. 2023;12(1):18.

38. San Miguel JF, Vidriales MB, Ocio E, et al. Immunophenotypic analysis of Waldenstrom's macroglobulinemia. *Semin Oncol*. 2003;30(2): 187-195.

39. Kriangkum J, Taylor BJ, Treon SP, Mant MJ, Belch AR, Pilarski LM. Clonotypic IgM V/D/J sequence analysis in Waldenstrom macroglobulinemia suggests an unusual B-cell origin and an expansion of polyclonal B cells in peripheral blood. *Blood*. 2004;104(7):2134-2142.

40. Sahota SS, Babbage G, Weston-Bell NJ. CD27 in defining memory B-cell origins in Waldenström's macroglobulinemia. *Clin Lymphoma Myeloma*. 2009;9(1):33-35.

41. Rinkenberger JL, Wallin JJ, Johnson KW, Koshland ME. An interleukin-2 signal relieves BSAP (Pax5)-mediated repression of the immunoglobulin J chain gene. *Immunity*. 1996;5(4):377-386.

42. Rao S, Karray S, Gackstetter ER, Koshland ME. Myocyte enhancer factor-related B-MEF2 is developmentally expressed in B cells and regulates the immunoglobulin J chain promoter. *J Biol Chem*. 1998;273(40):26123-26129.

43. Lin KI, Angelin-Duclos C, Kuo TC, Calame K. Blimp-1-dependent repression of Pax-5 is required for differentiation of B cells to immunoglobulin M-secreting plasma cells. *Mol Cell Biol*. 2002;22(13): 4771-4780.

44. Kownatzki E. Reassociation of IgM subunits in the presence and absence of J chain. *Immunol Commun*. 1973;2(1):105-113.

45. Eskeland T, Christensen TB. IgM molecules with and without J chain in serum and after purification, studied by ultracentrifugation, electrophoresis, and electron microscopy. *Scand J Immunol*. 1975;4(3): 217-228.

46. Davis AC, Shulman MJ. IgM—molecular requirements for its assembly and function. *Immunol Today*. 1989;10(4):118-128; 127-118.

47. Treon SP, Cao Y, Xu L, Yang G, Liu X, Hunter ZR. Somatic mutations in MYD88 and CXCR4 are determinants of clinical presentation and overall survival in Waldenström macroglobulinemia. *Blood*. 2014; 123(18):2791-2796.

48. Xu L, Hunter ZR, Tsakmaklis N, et al. Clonal architecture of CXCR4 WHIM-like mutations in Waldenström macroglobulinaemia. *Br J Haematol*. 2016;172(5):735-744.

49. Roccaro AM, Sacco A, Jimenez C, et al. C1013G/CXCR4 acts as a driver mutation of tumor progression and modulator of drug resistance in lymphoplasmacytic lymphoma. *Blood*. 2014;123(26): 4120-4131.

50. Hunter ZR, Xu L, Yang G, et al. Transcriptome sequencing reveals a profile that corresponds to genomic variants in Waldenström macroglobulinemia. *Blood*. 2016;128(6):827-838.

51. Kaiser LM, Harms M, Sauter D, et al. Targeting of CXCR4 by the naturally occurring CXCR4 antagonist EPI-X4 in Waldenström's macroglobulinemia. *Cancers*. 2021;13(4):826.