

The Evolution and Subtypes of Waldenstrom Macroglobulinemia: Findings from a Multi-omic Analysis of 249 Treatment Naive MYD88L265P Mutated Patients.

Zachary Hunter

`zachary_hunter@dfci.harvard.edu`

Dana-Farber Cancer Institute <https://orcid.org/0000-0002-1689-1691>

Maria Luisa Guerrero

Dana-Farber Cancer Institute

Nicholas Tsakmaklis

Dana-Farber Cancer Institute

Amanda Kofides

Dana-Farber Cancer Institute

Xia Liu

Dana Farber Cancer Institute

Lian Xu

Dana-Farber Cancer Institute

Guang Yang

Dana-Farber Cancer Institute

Joshua Gustine

Dana-Farber Cancer Institute

Andres Gamero

Dana-Farber Cancer Institute

Alberto Guijosa

Dana-Farber Cancer Institute

Shirong Liu

Dana-Farber Cancer Institute <https://orcid.org/0000-0002-3821-9134>

Hao Sun

Dana-Farber Cancer Institute

Kirsten Meid

Dana-Farber Cancer Institute

Christopher patterson

Dana Farber Cancer Institute

Mariateresa Fulciniti

Dana-Farber Cancer Institute <https://orcid.org/0000-0002-1432-9742>

Mehmet Samur

Dana Farber Cancer Institute <https://orcid.org/0000-0002-9978-5682>

Filip Garbicz

Dana-Farber Cancer Institute, Harvard Medical School <https://orcid.org/0000-0002-5548-1564>

Ruben Carrasco

DanaFarber Cancer Institute, Harvard Medical School

Andrew Branagan

Massachusetts General Hospital Cancer Center

Ari Melnick

Weill Cornell Medicine <https://orcid.org/0000-0002-8074-2287>

Shayna R. Sarosiek Sarosiek

Dana Farber Cancer Institute

Jorge Castillo

Dana-Farber Cancer Institute <https://orcid.org/0000-0001-9490-7532>

Kenneth Anderson

Dana Farber Cancer Institute <https://orcid.org/0000-0002-6418-0886>

Nikhil Munshi

Dana Farber Cancer Institute <https://orcid.org/0000-0002-7344-9795>

Steven TREON

Dana Farber Cancer Institute <https://orcid.org/0000-0001-6393-6154>

Article**Keywords:**

Posted Date: April 23rd, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-6473628/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

To study the heterogeneity of transcriptional and genomic traits that underlie the clinical presentation of untreated Waldenstrom's Macroglobulinemia (WM), we performed multi-omic studies in 249 treatment-naïve patients with WM. For all patients, RNASeq was performed on CD19-selected bone marrow (BM) MYD88 mutated samples as well as 13 paired CD19⁺CD27⁻ and CD19⁺CD27⁺ peripheral blood samples from healthy donors. Whole exome sequencing (WES) was also analyzed on paired tumor and germline DNA for 229 of these patients and DNA methylation status was assessed by enhanced reduced representational bisulfite sequencing (ERRBS) for 32 samples. The ERRBS and transcriptional data identified two distinct cells of origin for WM leading to two independent WM subtypes known as B-cell Like (BCL) and Plasma Cell Like (PCL). A third subtype, Early WM, was observed primarily in smoldering WM and was present in both ERRBS clusters identifying it as a transcriptional state common to early BCL and PCL. We further identified the WM Evolutionary Score (EScore), a gene signature common to all samples that predicted time to first therapy and corresponded with progression to symptomatic WM. Analysis of the EScore revealed a subclone expressing pre-B-cell, T-cell, stem cell and myeloid markers that was confirmed by flow cytometry and detected in early (smoldering) WM with subpopulations persisting with WM evolution. Our findings have important implications for WM pathogenesis. The evolution from the Early WM subtype to either BCL or PCL and EScore explains much of the genomic, transcriptional and clinical heterogeneity of WM including clinical outcomes to primary therapy.

Introduction

The past 15 years have been revolutionary for clinical and basic research in Waldenström's Macroglobulinemia (WM). Beginning with the initial descriptions of the highly recurrent activating mutations in *MYD88*^{1,2} and *CXCR4*³⁻⁶ found in 95% and 40% of WM patients, respectively, there have been multiple reports describing the associated mutational⁷⁻⁹, transcriptional¹⁰⁻¹², microenvironmental¹³⁻¹⁵, and epigenomic landscapes¹⁶⁻¹⁸ of WM. These discoveries have driven similar progress in the clinic leading to improved patient care as well as the development of BTK inhibitors (BTKi)¹⁹⁻²⁴ and BCL2 inhibitors²⁵ for WM. Despite this progress, the source of heterogeneity in clinical presentation and therapeutic outcomes remains largely unexplained as do the biological mechanisms that underlie progression from smoldering WM to symptomatic disease.

WM is an uncommon B-cell lymphoma meeting the clinicopathological definition of IgM secreting lymphoplasmacytic lymphoma characterized by accumulation of lymphoplasmacytic cells (LPC) predominately in the bone marrow (BM).²⁶ Like related lymphomas and multiple myeloma (MM), WM is thought to arise from monoclonal gammopathy of undetermined significance (MGUS) which progresses to active malignancy at a rate of ~ 1.5% per year.²⁷ The mechanisms of this transformation are not well understood and there is currently no consensus on whether IgM MGUS is a single entity or a handful of disease specific premalignant clones with similar clinical presentations.

To address these questions, we assembled the first large study of 249 untreated WM samples spanning from asymptomatic/newly diagnosed to patients being actively staged for therapy for multiomic analysis using whole exome sequencing (WES) and RNA sequencing (RNASeq).

Methods

Patient selection and sample processing

A review of WM patient samples collected at our clinic over the past 18 years identified 249 biopsies from untreated patients meeting the international consensus criteria for WM diagnosis.²⁶ Study samples represented the first visit to our clinic where WM BM samples were available. Samples were collected following informed consent in accordance with the Dana-Farber Cancer Institute (DFCI) Institutional Review Board. Mononuclear cells were isolated with gradient centrifugation using *Ficoll-Paque* (Amersham-Pharmacia Biotech, Piscataway, NJ) and WM LPCs were enriched with CD19⁺ MACS micro-bead selection (Miltenyi-Biotech, Auburn, CA). Germline DNA was isolated from paired peripheral blood samples following CD19 depletion. DNA and RNA were purified using the *AllPrep* mini kit (Qiagen, Valencia, CA). Fourteen paired healthy donor peripheral blood B-cells (HDPB; CD19⁺CD27⁻) and memory B-cells (HDMB; CD19⁺CD27⁺) were isolated using an immunomagnetic bead memory B-cell isolation kit (Miltenyi-Biotech). Samples were screened using our published allele-specific PCR assays for *MYD88* p.L265P and *CXCR4* p.S338X activating mutations with follow-up Sanger sequencing when possible.^{6,28} DNA from 5 IgM multiple myeloma (MM) and 8 CD138⁺ healthy donors plasma cells (HDPC) were provided by Dr. Nikhil Munshi (DFCI) for the methylation analysis. A detailed description of first-line therapies is available in the supplementary methods and summarized in Supplementary Tables 1-2.

Next Generation Sequencing

WES and RNASeq libraries were prepared by the Center for Cancer Computational Biology (DFCI). RNASeq was conducted on all samples using NEBNext Ultra II with poly-A selection and dUTP based strand specific construction (New England Biolabs, Ipswich, MA). For 229 samples, paired germline and tumor DNA were available for WES analysis using the HaloPlex Exome Target Enrichment System from Agilent Technologies (Santa Clara, CA). Libraries were sequenced on a *NextSeq 550* (Illumina, San Diego, CA) using 150 base paired-end reads. The RNASeq libraries were sequenced in partnership with Center for Cancer Genomic Discovery (DFCI) on a HiSeq2500 (Illumina) using 150 base paired-end reads. The enhanced reduced representational bisulfite sequencing (ERRBS) is part of a larger ongoing epigenomic study performed in partnership with the Epigenomics Core Facility of Weill Cornell Medicine using 50 base pair single reads using an Illumina HiSeq.²⁹ A repository for all sequencing data is in progress.

Flow Cytometry

BM and peripheral mononuclear cells were prospectively collected from WM patients and healthy donors for 13 color flow cytometric analysis using a LSR Fortessa (BD Biosciences, Franklin Lakes, NJ). A full

list of colors and clones used are shown in Supplemental Table 3. UltraComp eBeads Plus compensation beads (Thermo Fisher, Waltham, MA), Horizon Brilliant Stain Buffer Plus (BD Biosciences) and single staining controls were used on each run and the results analyzed using FlowJo (BD Biosciences).

Histological Analysis

For 66 samples, two slides cut from paraffin pathology blocks taken from the same biopsy as the study samples were stained with hematoxylin and eosin or Wright-Giemsa and digitally imaged for analysis.³⁰

Data Analysis

All sequencing data was aligned to GRCh38 using the Gencode v30 reference model. Analysis was performed in R (R Foundation for Statistical Computing, Vienna, Austria) with false discovery rate (FDR) applied to all appropriate results. A full description of analytical methods is available in the supplemental methodology.

Results

RNASeq was obtained for 249 treatment-naïve WM patients. For 229 of these patients, WES was obtained from the same biopsy. The clinical characteristics for the 249 patients were in line with previous reports.³¹ Their median age at time of BM biopsy was 65 (range 31.2 – 95) years; BM LPL disease involvement was 50% (4%-95%); serum IgM was 3,134 (104 – 9,274) mg/dl; 147 (59%) were male; 144 (58.5%) were symptomatic; 69 (27.7%) had adenopathy, and 21 (8.4%) splenomegaly. The median follow-up from the time from study BM biopsy was 7.2 (0-15.9) years and follow up from the date of diagnosis was 9.7 (0.7-33.8) years during which time 245 (98.4%) patients received at least one line of therapy and 49 (16.5%) died. The median time to first therapy from study BM biopsy was 0.3 (range 0-15.9) years.

Genomic features

All 249 patients carried the MYD88^{L265P} mutation and *CXCR4* mutations were identified in 113/249 (45.4%) patients by allele specific PCR and/or Sanger sequencing.^{6,28} For 229 patients who underwent WES, the following mutations were also identified as potential driver genes by dNdScv analysis³²: *ARID1A* (n=26; 11%), *BIRC3* (n=18; 8%), *CD79B* (n=14; 6%), *BTG2* (n=10; 4%), *H1-4* (n=9; 4%), *TRAF2* (n=7; 3%), *CTBP1* (n=7; 3%), *TRAF3* (n=6; 3%), *NDUFA7* (n=5; 2%) and *TNFAIP3* (n=2; 1%; Figure 1A). These findings were in line with previous reports.^{9,33,34} Deletions in chromosome 6q were also observed in 73/215 (34%) WES samples where copy number alterations (CNA) could be successfully called.

Multi-omic Feature Identification

Prior to this analysis, we noted rare cases where the *CXCR4* mutant gene signature was present in some samples without detectable *CXCR4* mutations and conversely missing in samples where *CXCR4*

mutations were confirmed in both DNA and mRNA.^{3,10} To assess if this signature was directly related to *CXCR4* mutation status or linked to a genetic background on which *CXCR4* mutations were likely to emerge, *CXCR4* mutant signature genes were used to cluster samples into two groups for differential gene expression (DGE) analysis using *CXCR4* mutation status as a covariate. The top 50 signature genes based on the principal component analysis (PCA) were used for hierarchical clustering identifying 3 distinct populations (Figure 1B; Supplemental Table 4). Based on gene set enrichment, the first two groups were identified as B-cell Like (BCL) and Plasma Cell Like (PCL; Figure 1C). A third group had intermediate levels of expression of all signature genes and was identified as Early WM (EWM) since this group had the longest time to first therapy ($p = 0.0016$; Figure 1D) and demonstrated the lowest rate of symptomatic disease with 31/75 (41%) being symptomatic, versus 64/103 (62%) and 49/68 (72%) with symptomatic disease for the BCL and PCL subtypes, respectively ($p < 0.01$ for both). This suggested that EWM subtyped patients may represent a common precursor state for both the BCL and PCL subtypes of WM. As predicted, the subtypes stratified *CXCR4* status with *CXCR4* mutations detected in 83/104 (79.8%) and 5/65 (7.2%; $p < 0.0001$) of BCL and PCL subtyped WM patients, respectively.

We also performed integrative multi-omic analysis with the top 1,000 high variance genes from the 229 samples with WES data using non-negative matrix factorization (NMF) to decompose the data into 3 metagenes as suggested by cophenetic and silhouette analysis (Figure 1E; Supplementary Figure 1A). The first of these two meta-genes split our samples into the same BCL and PCL subtypes (Supplemental Figure 1B). The third metagene correlated with BM involvement and symptomatic disease. Plotting all three metagenes together suggests a model where early/smoldering WM is predominantly the EWM subtype that evolves into either BCL or PCL as the disease progresses to symptomatic WM (Figure 1F). To further explore this shared evolutionary signature, the same 1,000 genes were analyzed using diffusion pseudotime (DPT) in the full data set (Figure 2A). EWM subtyped samples tightly clustered in early DPT and split into BCL and PCL branches as DPT progressed. This same pattern can be observed using PCA, though only when all the top three components were included (Supplementary Figure 1C). DPT, which we now term the WM Evolutionary Score (EScore), was strongly correlated with the first principal component, the first diffusion component, and metagene 3 (absolute r value > 0.94 ; $p < 0.0001$; Supplemental Figure 1D). Using a natural breakpoint in the DPT data space, samples were divided into Early and Late EScore (Supplemental Figure 1E). Early and Late EScore stratified time from biopsy to first therapy but not time from WM diagnosis to first therapy (Figure 2B-C). This same result was observed in multivariate Cox proportional hazard model (Figure 2D) demonstrating that EScore corresponds to progressing disease rather than an innate prognostic score. An example of independent interaction between EScore and subtype is seen in BM involvement. BCL and PCL subtypes showed different BM infiltration - 40% (5-90%) for BCL and 70% (15-95%) for PCL ($p < 0.0001$). These differences persisted across the EScore spectrum, with BM involvement of 30% (4-95%) in Early EScore and 70% (5-95%) in Late EScore patients ($p < 0.0001$; Figure 2E). As these categories are based on biological signaling, not clinical presentation, we sought to evaluate their correlation with clinical practice by classifying study marrows as either pre-therapy staging biopsies (≤ 12 weeks from start of treatment) or non-staging biopsies (> 12 weeks). EWM subtyped samples made up 51/115 (44.3%) and 24/130 (18.5%; $p < 0.0001$) while Early

EScore comprised 82/155 (72.2%) and 62/130 (47.7%; $p=0.0002$; Figure 2F) of non-staging, and staging BM biopsies, respectively.

WM Subtype

The top results from the DGE analysis comparing BCL to PCL are available in the Supplementary Data. WM subtype signature genes remained stratified by *CXCR4* mutation status, but visualizing *CXCR4* and subtype together clarified that the signal was driven by WM subtype (Supplementary Figure 2A). Many of the genes in this signature are not expressed by healthy donor B-cells or by any normal hematopoietic cell type. Even when these genes are ultimately silenced in one of the mature subtypes, they have robust expression in nearly all EWM subtyped WM samples suggesting dysregulated open chromatin may be a common feature of early WM disease (Figure 3A).

In addition to the BM findings, there were some notable trends from the clinical/pathological workup including expression of CD10 in 3/104 (2.9%) and 8/66 (21.1%), CD23 19/104 (38.5%) and 14/66 (21.2%), and CD5 19/103 (18.4%) and 4/66 (6.1%; $p=0.081$ for all) in BCL and PCL subtypes, respectively. A differential rate of adenopathy was also noted, but multivariate testing strongly suggests that this is linked to *CXCR4* mutation status. A complete list of BCL and PCL differences is available in Supplemental Table 5. In addition to *CXCR4*, *RALGAP1* mutations were associated with BCL while *EP300*, *IGHV3-20* and *IGHV3-23* mutations were indicative of PCL (Figure 3B). Trends in light chain hypermutation were also noted in BCL suggesting there may be some differences in immunoglobulin VDJ selection and hypermutation patterns between the subtypes. The *EP300* result together with a trend in *NOTCH1* with 1/82 (1.2%) in BCL and 6/63 (9.5%; $p=0.215$) in PCL, led us to investigate differences in NOTCH pathway mutations.³⁵ At least one NOTCH pathway gene was mutated in 20/82 (24.4%) and 27/63 (42.9%; $p=0.021$) of BCL and PCL patients, respectively.

Similar trends were observed in somatic CNAs with 6p amplifications in 3/77 (3.9%) and 11/62 (17.7%), 11q deletions in 1/77 (1.3%) and 9/62 (14.5%), 17p deletions in 0/77 (0%) and 6/62 (9.7%) and 18q amplifications in 11/77 (14%) and 1/62 (1.6%) in BCL and PCL, respectively ($p=0.054$ for all). While a trend was observed in 6q deletion rates between the subtypes with 23/77 (29.9%) in BCL and 29/62 (46.8%; $p=0.189$) in PCL, DGE analysis of the 6q deletion within each subtype demonstrated positional gene set enrichment of the whole arm in PCL while BCL was limited to 6q16-25 (Supplemental Figure 2B). The mean copy ratio across all 6q deleted samples from BCL and PCL respectively, confirmed this finding (Figure 3C; Supplemental Figures 2C-D).

ERRBS methylation data was available for 32 of the study samples along with 6 HDPB, 6 HDMB, 8 HDPC, and 5 IgM MM samples. A UMAP of high-quality methylation sites revealed two distinct clusters of WM with BCL clustering with HDMB and PCL clustering with IgM MM while HDPC and HDPB formed their own unrelated clusters (Figure 3D). As background CpG methylation status is highly correlated with cell of origin, this strongly suggests that the two subtypes arise from distinct populations, one of which is a memory B-cell and the other similar to IgM MM which is thought to originate from a germinal center B-

cell.^{36–38} This may explain the different expression and mutation patterns observed between the two subtypes in spite of having the same initiating event in *MYD88*. The EWM samples were found in both clusters suggesting that this group is a shared transcriptional phenotype associated with Early WM that gives way to subtyped expression as the clone evolves.

Plotting a UMAP of the top 500 high variance genes in *MYD88* mutated WM revealed novel structures on either extreme of the evolutionary track (Figure 3E). By performing DGE analysis of these groups with their respective parent subtype we were able to extract signature genes for clustering and identification of similar samples, which were labeled extreme BCL (BCE) and extreme PCL (PCE), respectively. This expanded classification improved the stratification of B-cell and plasma cell genes such as *PAX5*, *PRDM1* and *XBP1* (Supplemental Figure 2E). PCE was the only subtype with meaningful *SDC1* (CD138) expression. The BCE group differed from BCL in several ways including higher BM involvement: 75% (35–90%) versus 35% (5–90%; $p=0.003$); lower hemoglobin: 10 g/dL (8.4–12.9 g/dL) versus 12.2 g/dL (7.5–16.8 g/dL; $p<0.003$) and more patients with symptomatic disease: 15/15 (100%) versus 49/88 (55.7%; $p<0.001$). While the pattern of smaller deletions in 6q was common to both groups, the rate of 6q deletions was higher in BCE subtyped patients with 9/14 (64.3%) versus 14/63 (22.2%) in BCL ($p=0.064$). A trend for more *ARID1A* mutations was also noted in BCE with 5/15 (33.3%) versus 5/67 (7.5%) in BCL ($p=0.217$).

The PCE group had higher levels of BM plasma cells 4%; (0–6%) versus 0% (0–<1%) in PCL ($p=0.001$). Trends that were significant prior to FDR adjustment included more males with 13/14 (92.9%) versus 30/55 (54%; $p=0.16$), 6p amplifications 5/12 (41.7%) versus 6/50 (12%; $p=0.26$) and 6q deletions in 10/12 (83.3%) versus 19/50 (38%; $p=0.15$) among PCE versus PCL subtyped patients, respectively. DGE results for BCE and PCE compared with their host subtypes are available in the Supplemental Data.

While neither the EScore or WM subtypes associated with differences in overall survival, a log-rank analysis of primary therapies revealed that subtype did impact progression free survival (PFS) for patients who received proteasome inhibitor-based therapies including bortezomib dexamethasone and rituximab ($p=0.019$) as well as carfilzomib, dexamethasone and rituximab combination therapy ($p=0.002$) with the PCL and PCE subtypes displaying inferior PFS. This same effect could be seen when combining all proteasome inhibitor-based therapeutic strategies together ($p=0.0006$; Figure 3F).

WM Evolutionary Score

To better characterize incremental changes in the WM EScore, it was divided into 5 progressive levels (Supplementary Figure 3A–B). With this model, we can observe stepwise progression of somatic mutation burden and WM BM involvement (Figure 4A–B). DGE analysis of Early and Late EScore revealed significant enrichment for the *Hallmark Inflammatory Response* and *Hallmark Epithelial Mesenchymal Transition* that decreased with EScore (Figure 4D). The downregulation of inflammatory genes with EScore was particularly pronounced and genes such as the S100 alarmins correlated tightly with EScore ($r = -0.91$; $p<0.0001$; Figure 4E). To see if this decrease in inflammatory signaling resulted in observable changes in BM histology, pathology slides from 66 blocks matching the study biopsies were grouped by

EScore and analyzed by an independent pathologist revealing increased immature myeloid cells and eosinophils clustering near small lymphoid aggregates at low EScore values which gave way to increased lymphocytes and mast cells as the EScore increased (Figure 5A). This correlated with our observations from the epithelial to mesenchymal transition (EMT) enrichment which highlighted expression of the myeloid chemoattractants *CXCL1*, *CXCL8* and *CXCL12* in Early EScore as well as explained significance of the top gene ontology (GO) enrichment results which corresponded to myeloid activation and degranulation (Figures 5B-C).³⁹

To check for novel non-linear associations with EScore, generalized additive models (GAM) were employed and filtered against our previous DGE results. This identified several notable genes including *IL17RA* and *MYD88*, both of which were significantly elevated in at low EScore levels (Figure 5D). It also revealed the upregulation of B-cell markers *CD19* and *MS4A1* (CD20) with increasing EScore matching the downregulation of *CD38* and *SDC1* (CD138) observed in our original DGE analysis (Figure 5E).

We previously described the expression of the VDJ recombination genes *RAG1*, *RAG2* and *DNTT* in a subset of WM samples.¹⁰ In this analysis, these genes along with *MYB* and the pre-B-cell receptor (pre-BCR) surrogate light chain genes *IGLL1* and *VPREB1* were robustly expressed and decline significantly as EScore increased (Figure 6A). Among these genes, *IGLL1* had the highest median in HD samples at 0.09 (0-0.927) TpM. Median expression in Early EScore WM was 7.4 (0.04-76.95), 8.47 (0.02-80.29), 64.24 (0.2-581.19), 13.71 (0.37-116.66), 33.42 (0.13-305.01), and 103.35 (0.98-960.64) TpM for *RAG1*, *RAG2*, *DNTT*, *MYB*, *VPREB1*, and *IGLL1*, respectively ($p < 0.0001$ for all). Other examples of immunophenotype gene dysregulation included expression of T-Cell associated genes *CD3E*, *CD8*, and *CD4* along with myeloid and NK cell markers *CD33*, *FCGR3A* (CD16), and *CD14* all of which decreased with EScore ($p < 0.0001$ for all; Figure 6B). Low levels of the stem cell markers *CD34*, *KIT*, and *PROM1* ($p < 0.0001$ for all) were also observed. While *XBP1* and *PRDM1* varied by subtype, they were invariant with EScore, consistent with a model of dysregulated expression within a mature B-cell (Figure 6C). To verify this anomalous expression, WM and HD samples were prospectively collected for 13-color conventional flow cytometry analysis. Upon exclusion of doublets and dead cells, CD19⁺ Kappa⁺ or Lambda⁺ light chain cells were analyzed based on the established tumor clonality (Figure 6D). The assay was optimized using mononuclear cells from 4 WM patients and 1 HD. These studies confirmed the protein level expression of two or more of the non-B lineage markers CD34, CD16, CEACAM8 (CD66b), CD33, and CD4 in WM samples screened.⁴⁰ The fully optimized assay was then run on one additional untreated and one relapsing patient post ibrutinib and venetoclax combination therapy and 3 healthy donors (Supplementary Figure 4-5). Clinical characteristics of all 6 patients analyzed are shown in Supplementary Table 4.

Patients with Late EScore trended for improved PFS following BTK inhibitor-based (BTKi) therapy ($p = 0.098$; Figure 7A). Further analysis revealed that no patient in the Very Late EScore category had progressed on ibrutinib ($p = 0.035$) or within the broader BTKi category ($p = 0.026$; Figure 7B), respectively. Mirroring the increased B-cell immunophenotype expression, Late EScore patients demonstrated

enrichment for BCR signaling genes (Figure 7C; $p=0.007$) which may indicate increased reliance on MYD88/BCR signaling upstream of BTK in advanced WM. This finding contrasted with proteasome inhibitor-based therapies where there was no relationship between Early and Late EScore ($p=0.24$), though Very Late EScore was associated with inferior PFS despite being equally distributed between the BCL and PCL subtypes of WM ($p=0.0001$; Figure 7D).

Discussion

The findings from this study represent the largest multi-omic analysis undertaken in treatment-naïve WM patients and suggests that for *MYD88* mutated WM, there are three subtypes: EWM, BCL, and PCL. EWM is more prevalent in asymptomatic (smoldering WM) patients and represents an early evolutionary stage for the BCL and PCL subtypes. Our findings further affirm distinct cells of origin for the BCL and PCL subtypes of WM and a shared evolutionary trajectory from MGUS to symptomatic WM. These distinctions explain much of the known transcriptional and mutational heterogeneity of WM and also impact clinical presentation and therapeutic outcomes (Fig. 7E). The differences in CNA and mutational patterns between BCL and PCL suggests important differences in the underlying biology and may lead to divergent evolutionary patterns in response to therapy over time. While this is a novel framework for understanding WM biology, it is well-supported by previous work by Oaks and colleagues who previously described a memory B-cell like and plasma cell like clusters in their methylation analysis of WM.¹⁷ More recently, single-cell studies have identified a *DUSP22* high expression group with plasmacytic differentiation that corresponds to the PCL subtype.^{11,41}

Our findings are the first description of the EWM subtype which is predominant in smoldering WM patients, and which represents a transient transcriptional signature common to both BCL and PCL subtypes of WM. The finding that all genes associated with WM subtypes are expressed in the EWM subtype along with EWM clustering with BCL or PCL samples in the ERRBS data suggests that aberrantly open chromatin and lineage infidelity is likely a feature of early, evolving WM. This is in line with the multilineage disorganized expression observed by our studies in the early stages of WM development. While our study is the first to report this observation, supporting evidence can be found in the existing literature. The strong expression of S100A8 and S100A9 alarmins in MGUS and subsequent downregulation in WM was observed by Sklavenitis-Pistofidis and colleagues.¹¹ The expression of *RAG1*, *RAG2*, *MYB* and other pre/pro B-cell genes in a subset of patients was previously noted by us and others.^{10,42,43} Clonal populations of CD19⁺CD3⁺ B-cells have also been associated with stemness and early WM in several studies.^{41,44,45} The evidence presented here suggests WM arises from a memory B-cell or germinal center B-cell that has undergone an oncogenic transformation that reactivates a subset of stem cell programming leading to disorganized multilineage expression. This is consistent with the monoclonal IgM paraprotein that does not typically change during the disease course and the unchanging expression of *XBP1* and *PRDM1* throughout disease evolution. It does not rule out the possibility of a pre-neoplastic progenitor clone harboring the *MYD88* p.L265P mutation as suggested in some studies.^{46,47}

The discovery of the WM EScore is significant in several ways. It allows WM samples to be assessed by evolutionary stage permitting changes in BM histology to be seen as a function of progression rather than disease heterogeneity. Unlike the more clinically oriented asymptomatic WM risk score⁴⁸, the EScore provides valuable insights into the disease biology. The high expression of the alarmins, *CXCL8*, and *CXCL12* in early WM is notable as these are all associated with the formation of myeloid-derived suppressor cells that are known to be active in WM⁴⁹ and contribute to immune cell exhaustion, while also providing autocrine signaling through the Toll-like receptors and *CXCR4*. While the predominance of the multilineage gene expression in Early EScore suggests it represents a smoldering WM clone, its persistence through all disease stages and relapse may indicate that these may act as a cancer stem cell and/or serve as a clonal reserve driving progression. The strong expression of CD16 in the relapsing patient by flow cytometry is suggestive of such a mechanism. Studies into the therapeutic sensitivities of these cells and their role in WM pathogenesis are warranted. Likewise, the WM EScore permits insights into the biology of disease progression and therefore the ability to develop novel therapeutics aimed at preventing progression. Ultimately, these findings, especially those related to therapeutic outcomes should be independently confirmed. Tools for routine clinical assessment of the WM EScore and WM subtype are currently in development in our laboratory.

Declarations

Author Contributions:

NT, AK, XL, SL, CJP and HS, processed samples and performed laboratory experiments. JG, ARG, AG, KM, SRS, ARB, JJC and SPT identified samples and performed the clinical data analysis. AM and ZRH performed the methylation analysis. FG assisted with the acquisition, processing and staining of the pathology blocks. RC performed the blinded histopathological analysis. MLG designed and performed the flow cytometry analysis. JJC, SRS, SPT, KCA, and NM provided guidance on the clinical trial analysis. KCA, MS and MF, provided samples and insight for the ERRBS HDPC and IgM MM samples. ZRH and SPT designed the study and wrote the paper. ZRH performed the bioinformatic data analysis.

Acknowledgements

This work was supported by grants from the International Waldenstrom's Macroglobulinemia Foundation (ZRH, SPT); the Leukemia and Lymphoma Society Translational Research Grant (SPT); the Peter and Helen Bing Fund for WM Research; the Oszag and Kitchen WM Research Fund; the Paul and Ronnie fund for WM Research at the DFCI, the Estate of David and Barbara Levine, and the Dordick Law College. Dr. Hunter is supported by the Bliss Family Investigatorship in WM. Dr. Garbicz is supported by the Robert A. Kyle Career Development Award from IWMF. The authors would like to acknowledge John Quakenbush and Yaoyu Wang of the Center for Cancer Computational Biology for their assistance with the WES sequencing and library preparation, Arron Thorner, Anwesha Nag and the Center for Cancer Genomics for sequencing the RNASeq libraries, The Epigenomics Core Facility of Weill Cornell for performing the

ERRBS, Susan Lazo of the Dana-Farber Flow Cytometry Core for her help with the 13 color panel design, and The Brigham and Women's Specialized Histopathology Core for the slide staining and imaging. We would also like to thank all the patients who provided the samples to make this study possible.

References

1. Treon, S. P. *et al.* MYD88 L265P Somatic Mutation in Waldenström's Macroglobulinemia. *New England Journal of Medicine* **367**, 826–833 (2012).
2. Treon, S. P. & Hunter, Z. R. A new era for Waldenstrom macroglobulinemia: MYD88 L265P. *Blood* **121**, 4434–4436 (2013).
3. Treon, S. P. *et al.* Somatic mutations in MYD88 and CXCR4 are determinants of clinical presentation and overall survival in Waldenström macroglobulinemia. *Blood* **123**, 2791–2796 (2014).
4. Cao, Y. *et al.* The WHIM-like CXCR4S338X 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).
5. Hunter, Z. R. *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* **123**, 1637–1646 (2014).
6. Xu, L. *et al.* Clonal architecture of <sc>CXCR</sc> 4 <sc>WHIM</sc> -like mutations in Waldenström Macroglobulinaemia. *Br J Haematol* **172**, 735–744 (2016).
7. Hunter, Z. R. *et al.* Insights into the genomic landscape of MYD88 wild-type Waldenström macroglobulinemia. *Blood Adv* **2**, 2937–2946 (2018).
8. MacLachlan, K. H. *et al.* Waldenström macroglobulinemia whole genome reveals prolonged germinal center activity and late copy number aberrations. *Blood Adv* **7**, 971–981 (2023).
9. Varettoni, M. *et al.* Pattern of somatic mutations in patients with Waldenström macroglobulinemia or IgM monoclonal gammopathy of undetermined significance. *Haematologica* **102**, 2077–2085 (2017).
10. Hunter, Z. R. *et al.* Transcriptome sequencing reveals a profile that corresponds to genomic variants in Waldenström macroglobulinemia. *Blood* **128**, 827–838 (2016).
11. Sklavenitis-Pistofidis, R. *et al.* Single-cell RNA sequencing defines distinct disease subtypes and reveals hypo-responsiveness to interferon in asymptomatic Waldenstrom's Macroglobulinemia. *Nat Commun* **16**, 1480 (2025).
12. Bagratuni, T. *et al.* Single-cell analysis of MYD88^{L265P} and MYD88^{WT} Waldenström macroglobulinemia patients. *Hemasphere* **8**, e27 (2024).
13. Jalali, S. & Ansell, S. M. The Bone Marrow Microenvironment in Waldenström Macroglobulinemia. *Hematol Oncol Clin North Am* **32**, 777–786 (2018).
14. Vos, J. M. *et al.* CXCL13 levels are elevated in patients with Waldenström macroglobulinemia, and are predictive of major response to ibrutinib. *Haematologica* **102**, e452–e455 (2017).

15. Chen, J. G. *et al.* BTKCys481Ser drives ibrutinib resistance via ERK1/2 and protects BTKwild-type MYD88-mutated cells by a paracrine mechanism. *Blood* **131**, 2047–2059 (2018).
16. Chohan, K. *et al.* Multiomics analysis of IgM monoclonal gammopathies reveals epigenetic influence on oncogenesis via DNA methylation. *Blood* **144**, 1284–1289 (2024).
17. Roos-Weil, D. *et al.* Identification of two DNA methylation subtypes of Waldenström's macroglobulinemia with plasma and memory B cell features. *Blood* **136**, 585–595 (2020).
18. Hunter, Z. R. & Treon, S. P. Epigenomics in Waldenström macroglobulinemia. *Blood* **136**, 527–529 (2020).
19. Buske, C. *et al.* Ibrutinib Plus Rituximab Versus Placebo Plus Rituximab for Waldenström's Macroglobulinemia: Final Analysis From the Randomized Phase III INNOVATE Study. *Journal of Clinical Oncology* **40**, 52–62 (2022).
20. Tam, C. S. *et al.* A randomized phase 3 trial of zanubrutinib vs ibrutinib in symptomatic Waldenström macroglobulinemia: the ASPEN study. *Blood* **136**, 2038–2050 (2020).
21. Treon, S. P. *et al.* Ibrutinib in Previously Treated Waldenström's Macroglobulinemia. *New England Journal of Medicine* **372**, 1430–1440 (2015).
22. Xu, L. *et al.* Acquired mutations associated with ibrutinib resistance in Waldenström macroglobulinemia. *Blood* **129**, 2519–2525 (2017).
23. Treon, S. P. *et al.* Ibrutinib Monotherapy in Symptomatic, Treatment-Naïve Patients With Waldenström Macroglobulinemia. *Journal of Clinical Oncology* **36**, 2755–2761 (2018).
24. Treon, S. P., Xu, L. & Hunter, Z. MYD88 Mutations and Response to Ibrutinib in Waldenström's Macroglobulinemia. *New England Journal of Medicine* **373**, 584–586 (2015).
25. Castillo, J. J. *et al.* Venetoclax in Previously Treated Waldenström Macroglobulinemia. *Journal of Clinical Oncology* **40**, 63–71 (2022).
26. Treon, S. P. *et al.* Report of consensus Panel 4 from the 11th International Workshop on Waldenström's macroglobulinemia on diagnostic and response criteria. *Semin Hematol* **60**, 97–106 (2023).
27. Kyle, R. A. *et al.* Long-Term Follow-up of Monoclonal Gammopathy of Undetermined Significance. *New England Journal of Medicine* **378**, 241–249 (2018).
28. Xu, L. *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* **121**, 2051–2058 (2013).
29. Garrett-Bakelman, F. E. *et al.* Enhanced Reduced Representation Bisulfite Sequencing for Assessment of DNA Methylation at Base Pair Resolution. *Journal of Visualized Experiments* e52246 (2015) doi:10.3791/52246.
30. Shlomi, T., Segal, D., Ruppin, E. & Sharan, R. QPath: a method for querying pathways in a protein-protein interaction network. *BMC Bioinformatics* **7**, 199 (2006).
31. Treon, S. P. How I treat Waldenström macroglobulinemia. *Blood* **126**, 721–732 (2015).

32. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029-1041.e21 (2017).
33. Treon, S. P. *et al.* Genomic Landscape of Waldenström Macroglobulinemia and Its Impact on Treatment Strategies. *Journal of Clinical Oncology* **38**, 1198–1208 (2020).
34. Jimenez, C. *et al.* Genetic Characterization of Waldenstrom Macroglobulinemia By Next Generation Sequencing: An Analysis of Fouteen Genes in a Series of 61 Patients. *Blood* **126**, 2971–2971 (2015).
35. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371-385.e18 (2018).
36. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6 (2018).
37. Kuehl, W. M. & Bergsagel, P. L. Molecular pathogenesis of multiple myeloma and its premalignant precursor. *Journal of Clinical Investigation* **122**, 3456–3463 (2012).
38. Bolli, N., Martinelli, G. & Cerchione, C. The Molecular Pathogenesis of Multiple Myeloma. *Hematol Rep* **12**, 9054 (2020).
39. Cambier, S., Gouwy, M. & Proost, P. The chemokines CXCL8 and CXCL12: molecular and functional properties, role in disease and efforts towards pharmacological intervention. *Cell Mol Immunol* **20**, 217–251 (2023).
40. Guerrero, M. L. *et al.* Early Subclones Showing Aberrant Co-Expression of Stem Cell, T Cell and Myeloid Genes Can be Detected By Flow Cytometry in MYD88 Mutated Waldenström's Macroglobulinemia Patients. *Blood* **142**, 1952–1952 (2023).
41. Sun, H. *et al.* Single-cell profiles reveal tumor cell heterogeneity and immunosuppressive microenvironment in Waldenström macroglobulinemia. *J Transl Med* **20**, 576 (2022).
42. Gutierrez, N. C. *et al.* Gene Expression Profiling of B-Lymphocyte and Plasma Cell Populations from Waldenström's Macroglobulinemia. Comparison with Expression Patterns of the Same Cell-Counterparts from Other B-Cell Neoplasms. *Blood* **106**, 503–503 (2005).
43. Chng, W. J. *et al.* Gene-expression profiling of Waldenström macroglobulinemia reveals a phenotype more similar to chronic lymphocytic leukemia than multiple myeloma. *Blood* **108**, 2755–2763 (2006).
44. Qiu, Y. *et al.* Single-cell transcriptome analysis reveals stem cell-like subsets in the progression of Waldenström's macroglobulinemia. *Exp Hematol Oncol* **12**, 18 (2023).
45. Hao, M. *et al.* Gene Expression Profiling Reveals Aberrant T-cell Marker Expression on Tumor Cells of Waldenström's Macroglobulinemia. *Clinical Cancer Research* **25**, 201–209 (2019).
46. García-Sanz, R. *et al.* Clonal architecture and evolutionary history of Waldenström's macroglobulinemia at the single-cell level. *Dis Model Mech* **16**, dmm050227 (2023).
47. Rodriguez, S. *et al.* Preneoplastic somatic mutations including *MYD88*^{L265P} in lymphoplasmacytic lymphoma. *Sci Adv* **8**, eabl4644 (2022).

48. Bustoros, M. *et al.* Progression Risk Stratification of Asymptomatic Waldenström Macroglobulinemia. *Journal of Clinical Oncology* **37**, 1403–1411 (2019).

49. Bhardwaj, V. *et al.* Expanded tumor-associated polymorphonuclear myeloid-derived suppressor cells in Waldenstrom macroglobulinemia display immune suppressive activity. *Blood Cancer J* **14**, 217 (2024).

Figures

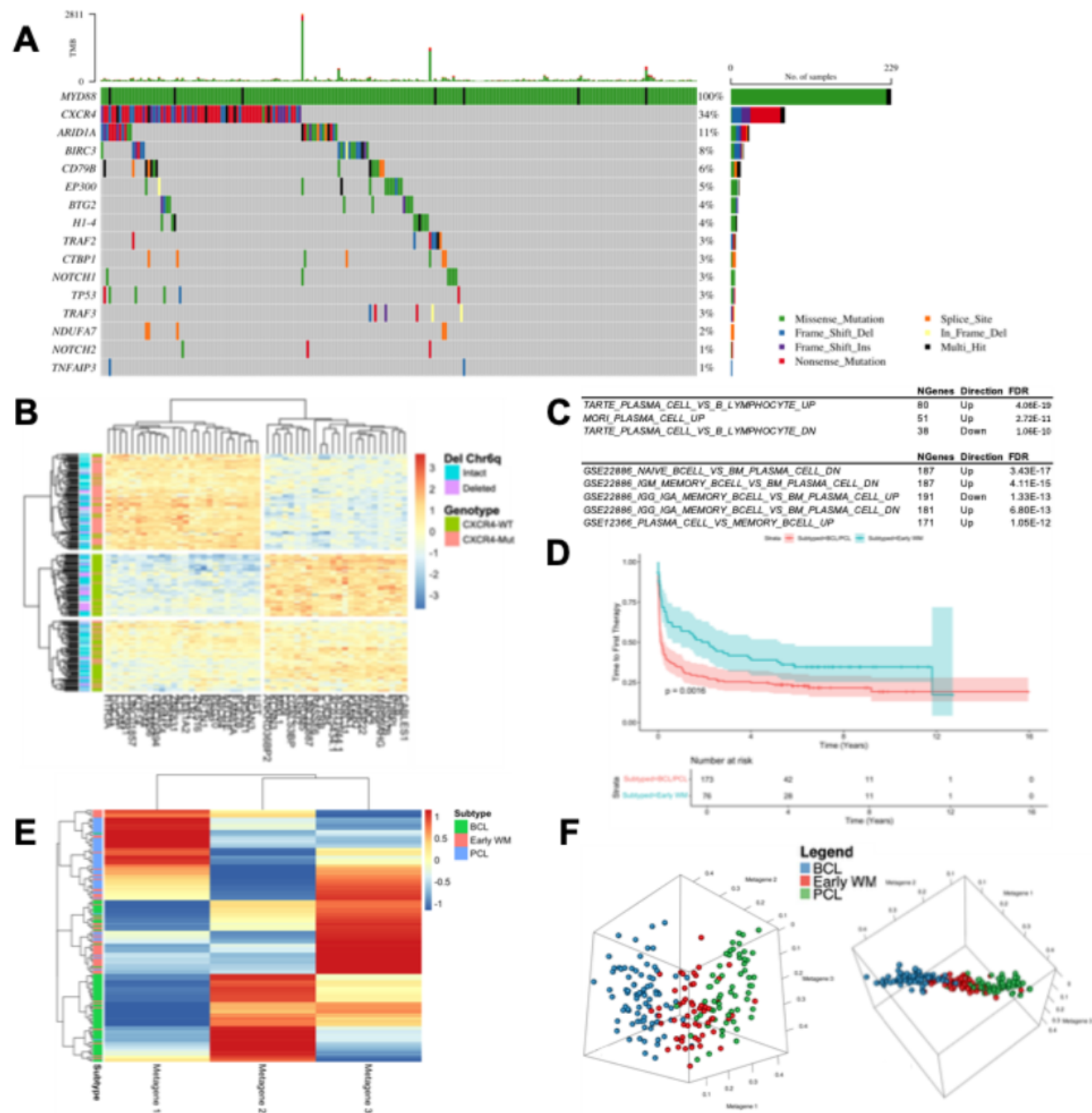


Figure 1

Identification of WM Subtypes. A) Oncoplot of driver genes identified by dNdScv and genes associated with WM. B) Heatmap of the top fifty genes based on the *CXCR4* mutant signature identified 3 clear expression clusters. From the top, these groups were characterized as Early WM (EWM); B-cell Like (BCL) or Plasma Cell Like (PCL). C) Top mSigDB Curated and Immunological gene set results contrasting BCL and PCL. D) Time to first therapy from biopsy for EWM, BCL or PCL subtypes of WM. E) Non-negative matrix factorization of top 1,000 high variance genes. F) Three-dimensional plotting of the NMF metagenes. The data was restricted to a plane in three-dimensional space suggesting that the need for 3 metagenes may be an artifact of the non-negative aspect of NMF and that a two-component model may be sufficient.

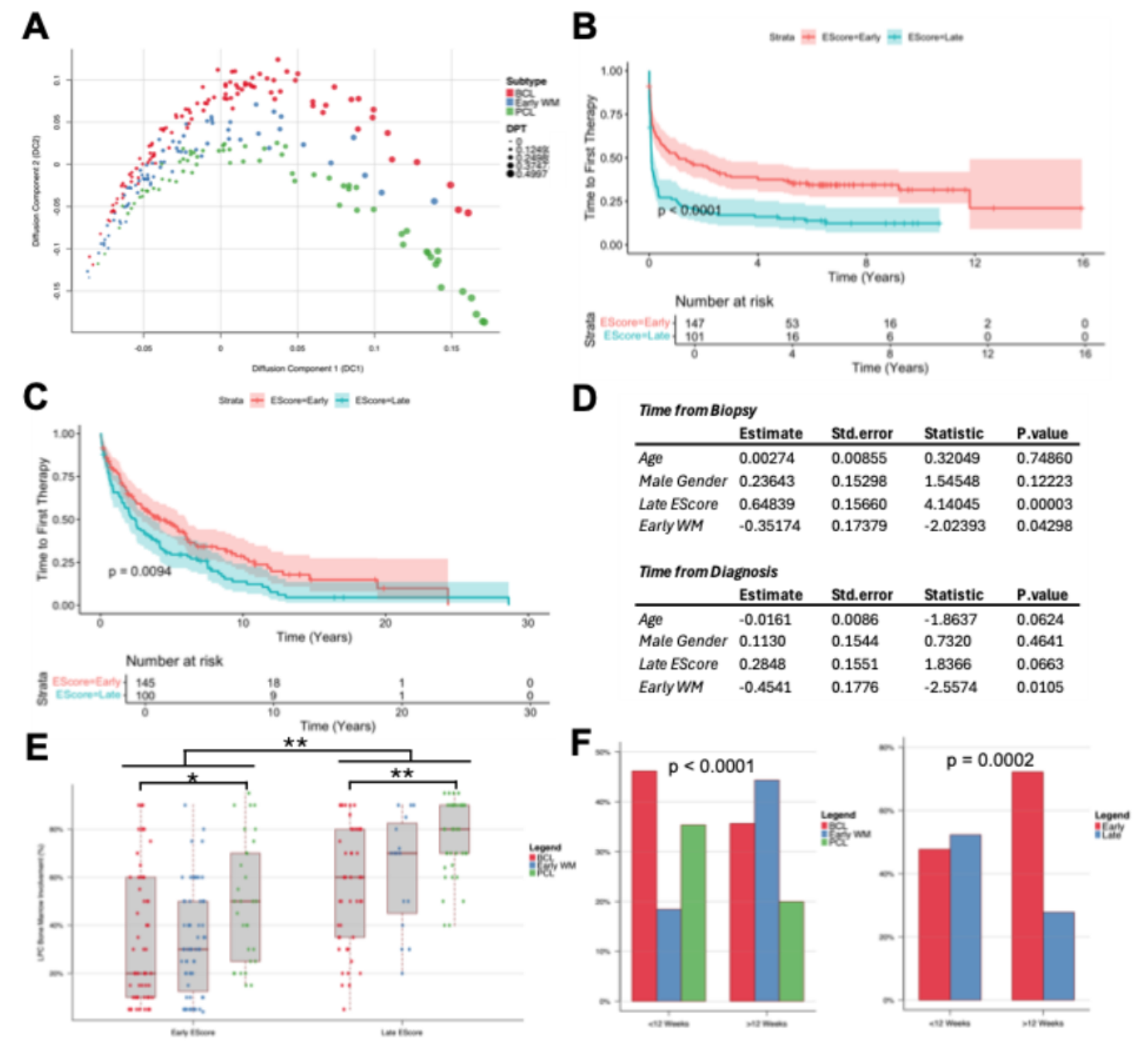


Figure 2

WM Evolutionary Score corresponds to progression of asymptomatic to symptomatic WM. A) Diffusion map colored by subtype and sized by DPT value. B) Early and Late EScore stratify time from study biopsy to first therapy. C) Time from WM diagnosis to first therapy stratified by Early and Late EScore. D) Cox proportional hazard models including age at biopsy, gender, Early/Late EScore and EWM status. Time from study biopsy to first therapy (top) and time from WM diagnosis to first therapy (bottom) were analyzed. E) Both EScore and WM subtype independently influence WM BM involvement. * $p=0.01$; ** $p<0.0005$. F) Percentage of patients categorized by subtype and EScore were stratified on timing of the study biopsy (<12 vs. >12 weeks) from the initiation of first therapy.

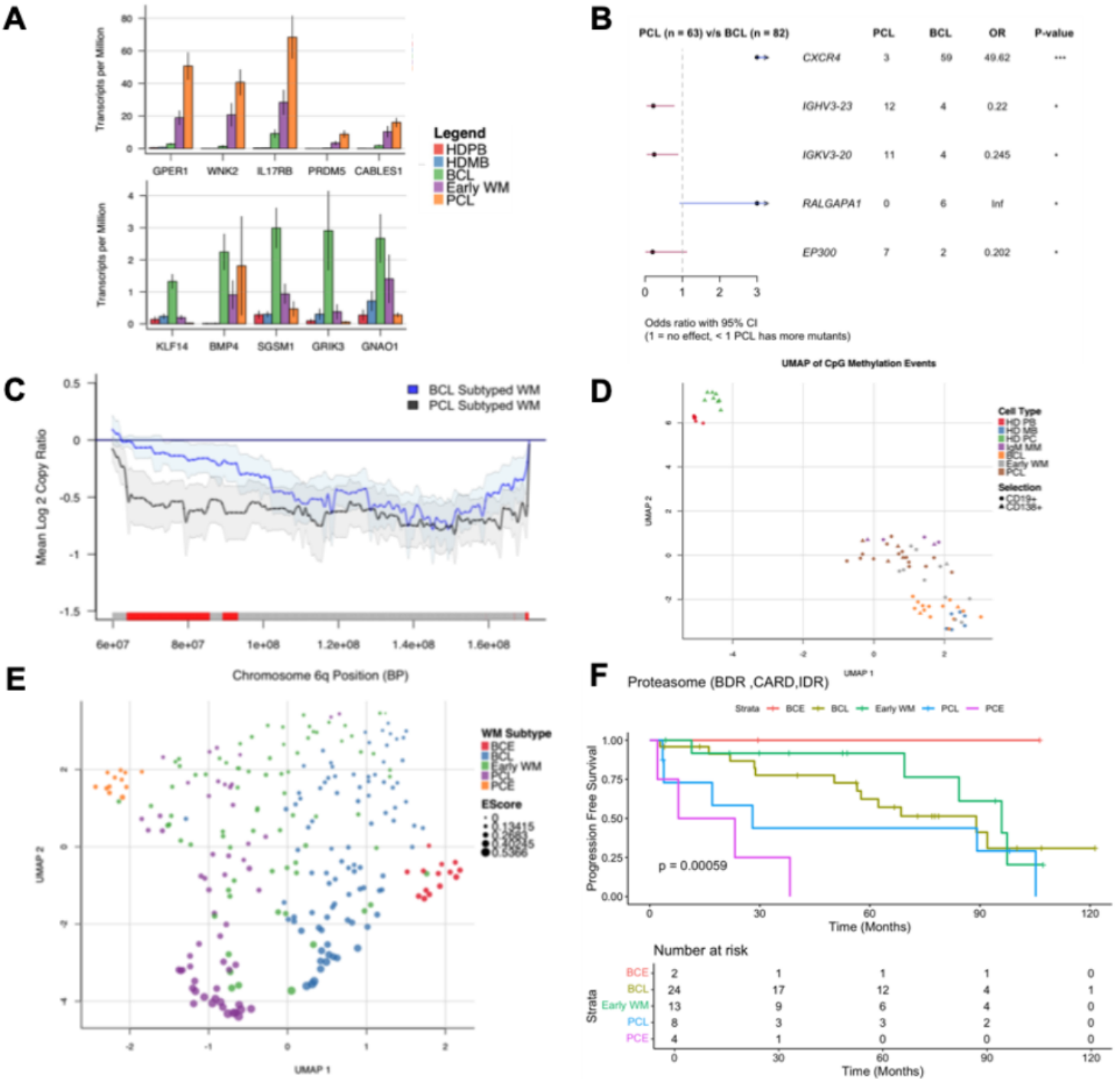


Figure 3

Characterization of the BCL and PCL subtypes. A) Dysregulated gene expression in PCL (top panel) and BCL (bottom panel). Many of these genes are not expressed in HD B-cells. EWM subtyped samples showed intermediate levels of expression of both BCL and PCL associated genes. B) Forest plot of mutated genes predictive of BCL versus PCL subtypes. * $p < 0.05$ *** $p < 0.0005$. C) Smoothed spline of mean log 2 copy ratios of patients with 6q deletions by WM Subtype. Shaded regions indicate 2x standard error of the mean. Red sections of the rug plot indicated FDR p -values < 0.05 . D) UMAP of high-quality CpG methylation data from the ERRBS analysis. E) UMAP of top 500 high variance genes in WM identifies the extreme BCL (BCE) and extreme PCL (PCE) subtypes which represent the extreme ends of the B-cell to plasma differentiation spectrum and have unique expression, genomic, and clinical characteristics. F) Progression-free survival analysis for WM patients receiving proteasome inhibitor-based therapy stratified by WM Subtype. The median study PFS was 52.6 months.

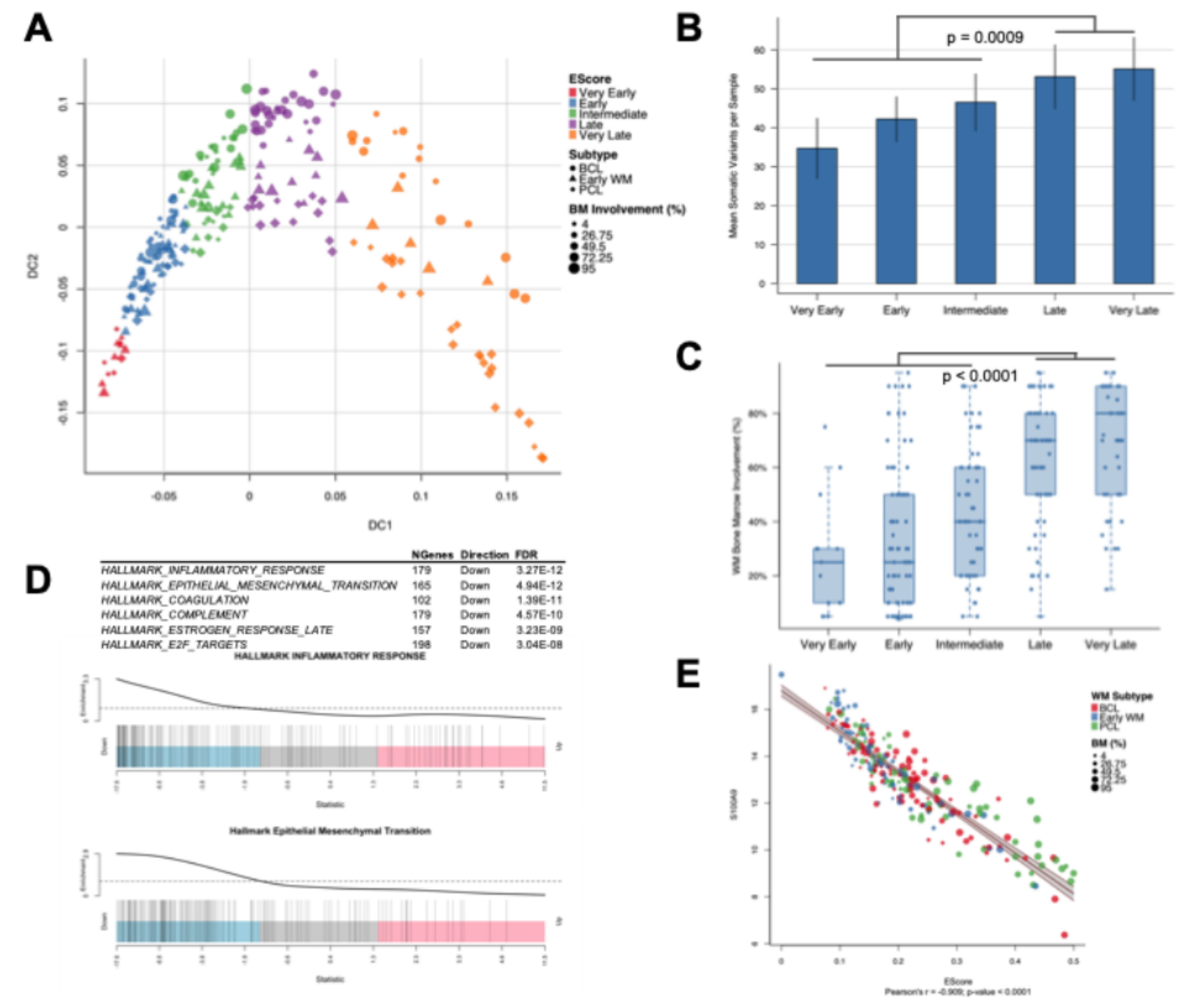


Figure 4

Characterization of the WM Evolutionary Score. A) Diffusion map of the WM study samples. Points are colored by EScore divided into 5 clusters: Very Early, Early and Intermediate (splitting the former “Early” EScore category), Late and Very Late (the latter two comprising the former “Late” category). Shapes represent WM subtype membership and are sized according to percent of WM intertrabecular involvement. B) Mean number of somatic variants per patient by EScore levels. To avoid skewing by outliers, samples with mutation loads outside of 3 times the inner quartile range (IQR) from the IQR boundaries were removed. Error bars represent 2 standard errors of the mean. C) WM intertrabecular involvement (%) by EScore. D) Hallmark gene set enrichment results from Late compared to Early EScore. E) Correlation of S100A9 with EScore. Data are colored by WM subtype and sized by BM involvement. The adjusted R^2 of the corresponding linear model was 0.826. Expression in all correlation and clustering models was represented by variance stabilizing transformation of the batch-adjusted count data.

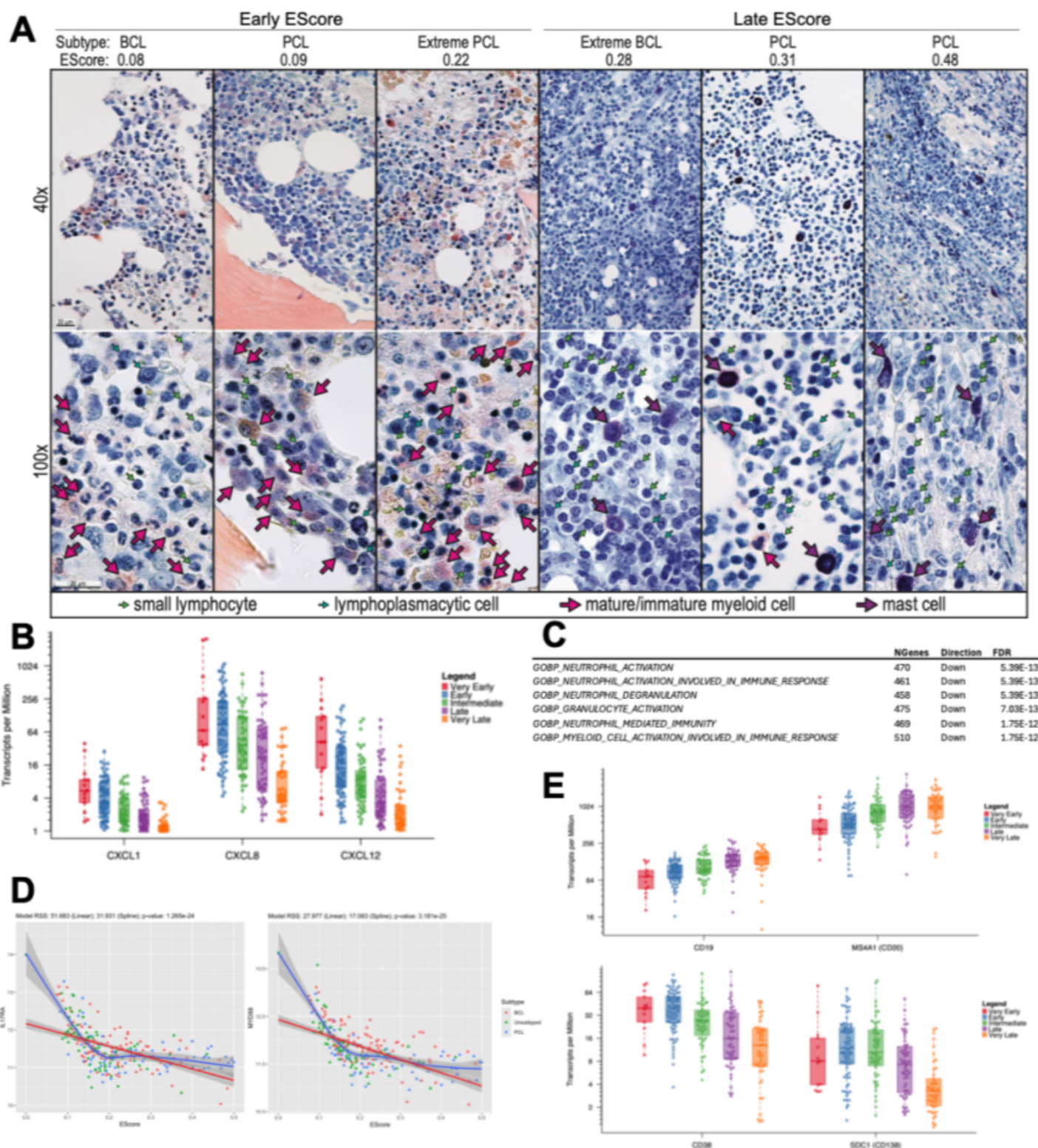


Figure 5

EScore impact on histology and immunophenotype. A) Representative Wright-Giemsa stained sections from the same biopsy as the study data ordered by progressive EScore. Early EScore samples demonstrated an abundance of immature myeloid precursors while sections from Late EScore samples showed large tumor sheets and mast cell involvement. B) Expression of myeloid chemo attractants *CXCL1*, *CXCL8*, and *CXCL12* were strongly associated with Early EScore. C) Top gene ontology gene set

enrichment results for Late vs. Early EScore. D) Generalized additive model testing identified nonlinear associations with EScore not detected in the DGE analysis including *IL17RA* (left) and *MYD88* (right). The blue line represents the fitted spline model compared with the linear model shown in red. E) GAM testing identified an increase of the B-cell immunophenotype markers *CD19* and *MS4A1* (CD20) with EScore as shown on top. Regular DGE testing identified the corresponding downregulation of *CD38* and *SDC1*(CD138) as shown beneath.

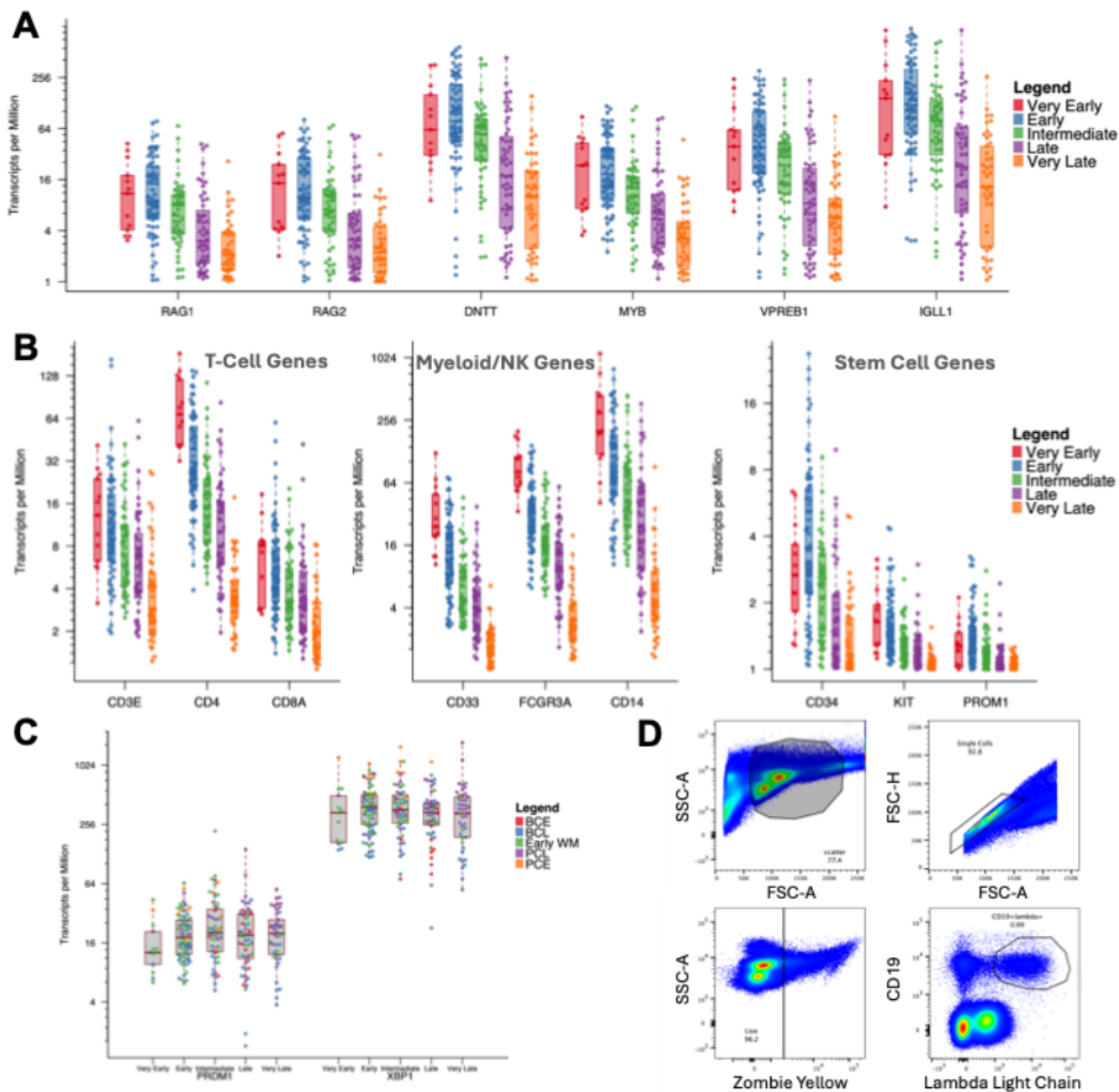


Figure 6

Identification and validation of a novel population of cells within the WM clone with disorganized expression from multiple cell lineages. A) Expression of pre/pro B-cell genes including VDJ related (RAG1, RAG2, DNTT), pre-BCR/surrogate light chain (VPREB1, IGLL1) and the early transcription factor/proto-oncogene MYB are highly expressed early in WM evolution (EScore). B) Expression grouped from the left of T-cell, Myeloid/NK, and stem cell genes by EScore level. C) Expression of key plasma cell transcription factors was unchanged with EScore consistent with a mature B-cell cell of origin. D) Gating strategy for 13 color flow within the CD19+ light chain restricted gate following exclusion of doublets and dead cells.

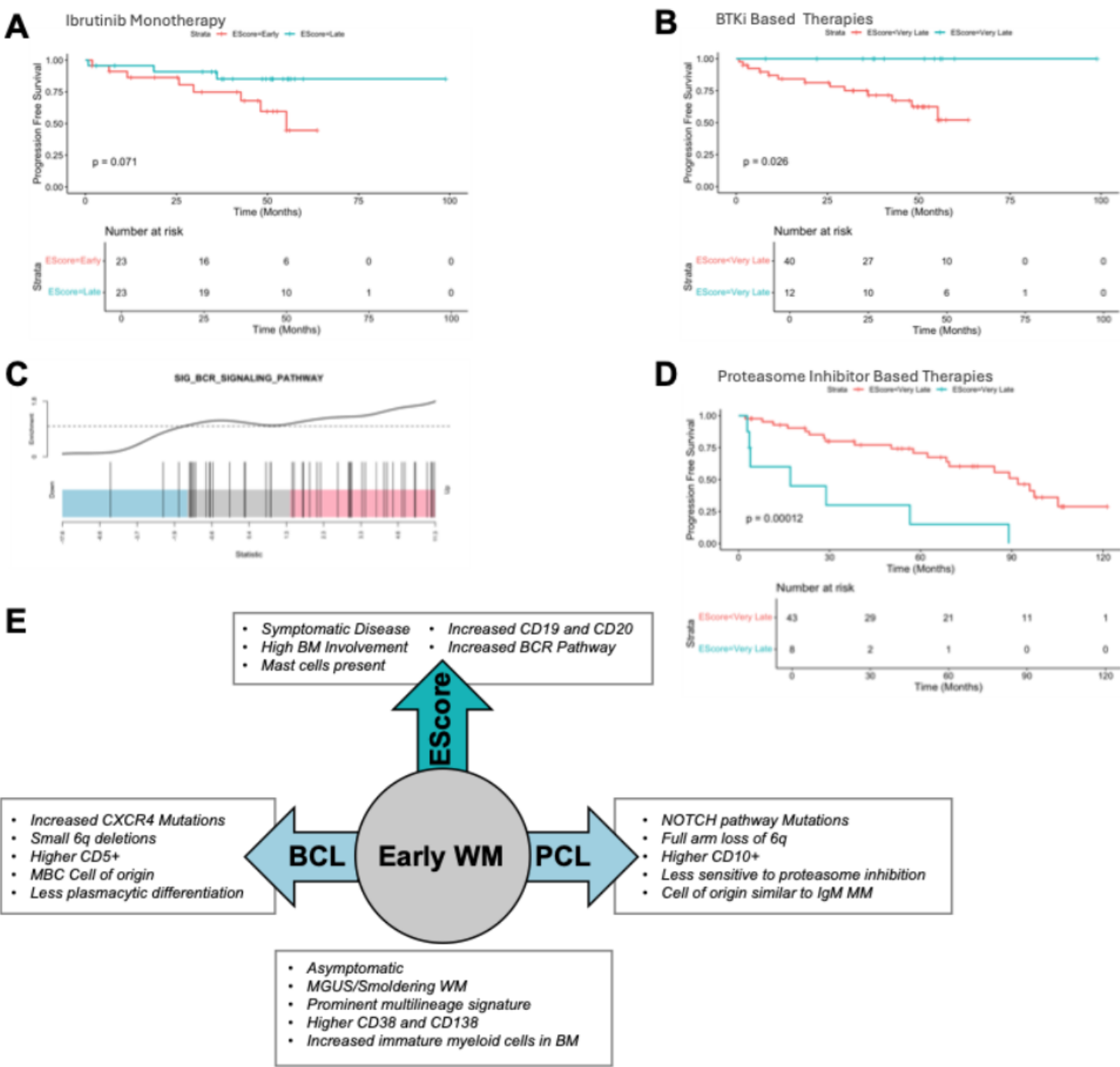


Figure 7

Clinical significance of the WM EScore. A) Progression free survival (PFS) for patients who received ibrutinib monotherapy stratified by Early and Late EScore. Median study follow up was 42.15 months. B) PFS for all BTK inhibitor (BTKi) based therapies for those with Very Late EScore compared to all others. Median study follow up was 38.2 months. This was significant for ibrutinib monotherapy as well with $p=0.035$. C) Gene set enrichment for genes within the B-cell receptor signaling pathway increasing with EScore suggesting increased reliance on MYD88/BCR signaling upstream of BTK. D) PFS for patients receiving proteasome-based therapy for patients with Very Late EScore compared with all others. Note that the Very Late group was split evenly between BCL and PCL subtypes. E) Summary of clinical and genomic impacts of WM subtype and EScore.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [MultiomicSupplementaryMethods41625.docx](#)
- [MultiomicsSupplementaryTables41625.docx](#)
- [MultiomicsSupplementaryFigures41625.docx](#)
- [DGEResults.xlsx](#)