REGULAR ARTICLE

blood advances

Integrated multiomic profiling reveals 2 distinct splenic marginal zone lymphoma subgroups with prognostic relevance

Helen Parker,[1,*] Amatta Mirandari,[1,*] Benjamin Stevens,[1,*] Carolina Jaramillo Oquendo,[2] Martí Duran-Ferrer,[3] Lara Buermann,[1] Harindra E. Amarasinghe,[1,4] Jaya Thomas,[1] Louise Carr,[1] Shama Syeda,[1] Methusha Sakthipakan,[1] Marina Parry,[1] Matthew J. J. Rose-Zerilli,[1] Zadie Davis,[5] Neil McIver-Brown,[5] Aliki Xochelli,[6] Sarah Ennis,[2] Lydia Scarfò,[7] Paolo Ghia,[7,8] Christina Kalpadakis,[9] Gerassimos Pangalis,[10] Davide Rossi,[11,12] Gianluca Gaidano,[12] Simon Wagner,[13] Matthew Ahearne,[13] Marc Seifert,[14] Christoph Plass,[15] Dieter Weichenhan,[15] Eva Kimby,[16] Lesley-Ann Sutton,[17] Richard Rosenquist,[17,18] Rose-Marie Amini,[19] Viktor Ljungström,[20] Guy Pratt,[21,22] Francesco Forconi,[1] Kostas Stamatopoulos,[6] Marta Salido,[23] Ana Ferrer,[23] Catherine Thieblemont,[24] Laura Hilton,[25] Ryan Morin,[25] Renata Walewska,[5] José Ignacio Martín-Subero,[3,26] David Oscier,[5] Christopher Oakes,[27] Jane Gibson,[1,28,†] Dean Bryant,[1,†] and Jonathan C. Strefford[1,28,†]

[1]Cancer Sciences, Faculty of Medicine, and [2]Genomic Informatics, Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom; [3]Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain; [4]Radcliffe Department of Investigative Medicine, University of Oxford, Oxford, United Kingdom; [5]Department of Molecular Pathology, University Hospitals Dorset, Bournemouth, United Kingdom; [6]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece; [7]Unit of B Cell Neoplasia and Strategic Research Program on Chronic Lymphocytic Leukaemia, Comprehensive Cancer Center, Instituto di Recovero e Cura a Carattere Scientifico Ospedale San Raffaele, Milan, Italy; [8]Medical School, Università Vita-Salute San Raffaele, Milan, Italy; [9]University Hospital, University of Crete, Heraklion, Greece; [10]Department of Haematology, Athens Medical Center, Athens, Greece; [11]Institute of Oncology, Università della Svizzera Italiana, Bellinzona, Switzerland; [12]Division of Hematology, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy; [13]Leicester Cancer Research Centre, Leicester Royal Infirmary, Leicester, United Kingdom; [14]Institut fur zellbiologie, Universitatsklinikum Essen, Essen, Germany; [15]Epigenomics and Cancer Risk Factors, German Cancer Research Center, Heidelberg, Germany; [16]Unit of Hematology, Department of Medicine Huddinge, and [17]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; [18]Clinical Genetics and Genomics, Karolinska University Hospital, Stockholm, Sweden; [19]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; [20]Department of Biomedical Informatics, Harvard Medical School, Boston, MA; [21]School of Cancer Studies, University of Birmingham, Birmingham, United Kingdom; [22]Department of Haematology, Heart of England NHS Foundation Trust, Birmingham, United Kingdom; [23]Laboratori de Citologia Hematològica i Citogenètica Molecular, Servei de Patologia, Hospital del Mar, Barcelona, Spain; [24]Service d'hémato-oncologie, Hôpital Saint-Louis, Université Paris Cité, Assistante Publique–Hôpitaux de Paris, Paris, France; [25]Department of Molecular Biology and Biochemistry, Simon Fraser Univesity, BC, Canada; [26]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain; [27]Division of Hematology, Department of Internal Medicine, The Ohio State University, Columbus, OH; and [28]Bioinformatics Research Facility, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

Key Points

- We identified 2 robust SMZL subgroups, with distinct methylation, mutational, chromosomal, immunogenetic, and transcriptional landscapes.
- The high-risk subgroup had poorer survival, a higher transformation risk, and longer treatment time, surpassing other molecular models.

Splenic marginal zone lymphoma (SMZL) is a rare B-cell malignancy with notable genetic, epigenetic, and clinical heterogeneity. In this study, we used coding and noncoding sequencing (n = 74), including whole-genome sequencing (WGS) of 24 paired tumor-normal samples, targeted sequencing (n = 55), and DNA methylation in 126 patients to characterize the disease. From WGS, we identified recurrent, predominantly clonal coding mutations in *KLF2* (50%), *KMT2D* (25%), and *NOTCH2* (25%), alongside rare mutations in *FLNC* (8%), novel mutations in *FAM135B* (17%), and noncoding mutational hot spots in *BCL6*, *PAX5*, and *BACH2* linked to aberrant somatic hypermutation. At least 1 noncoding hot spot was detected in 69% of patients. Copy number aberrations were present in 73% of patients, including del(7q) (27%), gain(3q) (17%), and trisomy 12 (13%). DNA methylation profiling revealed 2 epigenetic subgroups: high-risk (HR) SMZL (n = 67) and low-risk SMZL (n = 59). SMZL-HR was associated with adverse features, including female sex, IGHV1-2*04 usage, *KLF2* mutations, del(7q), shorter telomeres, and elevated epigenetically determined cumulative mitoses scores. Transcriptomic analysis highlighted enhanced cell proliferation in SMZL-HR, with enrichment of E2F and G2M checkpoint pathways and

Submitted 9 December 2025; accepted 16 April 2026; prepublished online on *Blood Advances* First Edition 27 May 2026; final version published online 16 July 2026. https://doi.org/10.1182/bloodadvances.2025019336.

*H.P., A.M., and B.S. are joint first authors.

†J.G., D.B., and J.C.S. are joint senior authors.

Methylation array data have been deposited in the BioStudies database (accession number E-MTAB-15791).

All other data supporting the findings of this study are available from the corresponding author, Jonathan C. Strefford (jcs@soton.ac.uk), on request.

The full-text version of this article contains a data supplement.

© 2026 American Society of Hematology. Published by Elsevier Inc. Licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), permitting only noncommercial, nonderivative use with attribution. All other rights reserved.

epigenetic regulation via *EZH2*. Patients with SMZL-HR had significantly shorter time to first treatment (TTFT) (hazard ratio, 1.9; $P = .003$) and reduced overall survival (hazard ratio, 2.5; $P = .039$): 85% of patients with SMZL-HR required treatment and showed a higher frequency of transformation ($P = .007$) and mortality ($P < .001$). Multivariate analysis confirmed SMZL-HR as an independent predictor of shorter TTFT (hazard ratio, 2.4; $P = .001$). These findings demonstrate the role of DNA methylation and molecular profiling in SMZL risk stratification.

Introduction

Splenic marginal zone lymphoma (SMZL) is a rare mature B-cell malignancy primarily involving the spleen, bone marrow, and peripheral blood.[1-4] SMZL typically pursues an indolent course, with a median survival of 10 to 15 years, although only 40% to 50% of deaths are disease- or treatment-related. However, ~30% of patients have shorter survival despite treatment, and prognosis is especially poor for 5% to 15% of patients that undergo transformation to large-cell lymphoma.[5,6]

Scoring systems based on routinely measured parameters, namely, hemoglobin, lactate dehydrogenase, albumin,[6] platelet count, and extrahilar lymphadenopathy,[7] can divide patients into risk groups with differing 5-year cause-specific or lymphoma-specific survival. Importantly, no difference in lymphoma-specific or progression-free survival was found between patients treated with rituximab and chlorambucil at diagnosis and those receiving the same treatment after a period of watchful waiting.[8] Consistent with these data, current indications for initiating treatment are based on clinical criteria, namely, progressive and/or symptomatic disease.

Biomarker studies in SMZL have identified disease-specific genomic, immunogenetic, and epigenetic features, providing insights into disease biology and prognosis that may have potential value in clinical trial design and as therapeutic targets. Their prognostic significance is exemplified by 2 seminal studies. Arribas et al[9] identified 2 clusters of patients with differing levels of promotor methylation, showing that higher promotor methylation was associated with use of the immunoglobulin heavy chain variable (IGHV) 1-02 gene, *NOTCH2* mutations, 7q31-32 loss, histologic transformation, and shorter overall survival (OS). Bonfiglio et al[10] described 2 major genetic clusters and 2 differing splenic immune microenvironments. Patients with mutations involving *NF-kB* genes or *KLF2* and an immunosuppressive microenvironment had significantly higher 10-year mortality than a matched general population.

We sought to extend these findings by performing the first whole-genome sequencing (WGS) analysis of tumor–germ line duos in SMZL and employing more recent DNA methylation arrays with genome-wide coverage, the latter of which permit analysis of epigenetically determined cumulative mitoses (epiCMIT), a DNA methylation-based mitotic clock that reflects the proliferative history of B-cell tumor samples.[11] Our study includes patients without treatment and integrates genetic, immunogenetic, phenotypic, transcriptomic, and epigenetic data to delineate more precise molecular subgroupings and their clinical implications.

Methods

Patient cohort

Our cohort included 133 patients with SMZL diagnosed between 1992 and 2021 at centers with expertise in the disease who met established diagnostic criteria (Table 1; Figure 1A). Informed consent was obtained from all patients in accordance with the Declaration of Helsinki, and the study was approved by the regional and local ethics committees of participating institutions. Patients were sampled before treatment. Before DNA extraction (DNeasy Blood and Tissue Kit, Qiagen, Hilden, Germany), SMZL cells were purified from peripheral blood (n = 116) or spleen (n = 17) using the EasySep Human B Cell Enrichment Kit without CD43 depletion (Stemcell Technologies, Cambridge, United Kingdom). Tumor purity >85% was confirmed in all patients by flow cytometry ($CD19^+/CD45^+$) (supplemental Figure 1).

WGS

WGS was performed on 24 unselected SMZL patients (termed the WGS cohort) using paired tumor B cells and saliva (DNeasy Blood and Tissue Kit, Qiagen, and Oragene DNA Saliva Kit, DNA Genotek, Ottawa, Canada) with the SNP&SEQ Technology Platform (Science for Life Laboratory, Uppsala University, Sweden). Data was analyzed using custom pipelines based on Genome Analysis ToolKit (GATK) Best Practices workflows for somatic variant calling (supplemental Methods). Coding tumor mutational burden (TMB) per megabase and kataegis regions were identified using MAFTools version 2.10.0. Mutational signatures were analyzed using SigProfilerExtractor (version 1.1.4) and SigMiner (version 2.1.3) and compared with the Catalogue of Somatic Mutations in Cancer.[12] Copy number alterations (CNAs) were called from WGS data using Battenberg (version 2.2.10), and PhyloWGS (version 1.0-rc1) was used to infer phylogenetic trees and determine the subclonal composition and evolution of individual samples (supplemental Methods). Structural variants (SVs) were called using Delly (version 1.0.3). Each SV required at least 5 reads covering each breakpoint. Only balanced rearrangements or those resulting in concomitant CNA were considered. Each SV was manually curated by 2 experts.

Targeted resequencing

An additional 105 patients were screened using 2 HaloPlex resequencing panels (Agilent Technologies, Santa Clara, CA). Both panels targeted 45 genes/genomic loci selected for their established relevance in SMZL and other mature B-cell tumors, as previously described.[13] The second panel also captured 15 413 novel coding and noncoding genes/genomic loci based on our WGS data and was applied to 50 patients (termed the WGS extended

Table 1. Key clinicobiological characteristics of the SMZL cohort

Characteristics	Patients, % (n/N)
Female	59 (78/133)
Age at diagnosis, median (range), y	68 (36-86)
Patients treated, including splenectomy	72 (95/132)
Patients transformed to large B-cell lymphoma	26 (17/66)
Patients who died	27 (36/131)
OS, median (range), mo	63 (3-293)
TTFT, median (range), mo	13 (0-179)
HPLL/ABC score	
A	78 (54/69)
B/C	22 (15/69)
IGHV gene usage	
IGHV1-2*04	25 (27/107)
IGHV 3-23	15 (16/107)
IGHV 4-34	6 (6/107)
IGHV 3-30	9 (10/107)
IGHV germ line identity	
100%	9 (8/94)
97%-99.9%	39 (37/194)
<97%	52 (49/94)
Gene mutations	
ARID1A	5 (6/129)
FLNC	5 (7/129)
KLF2	29 (38/129)
KMT2D	21 (27/129)
MYD88	5 (7/129)
NOTCH2	21 (27/129)
TNFAIP3	13 (17/129)
TP53	16 (20/129)
TRAF3	15 (19/129)
Bonfiglio NNK-like	29 (78/129)
Cytogenic abnormalities	
Gain 3q	17 (21/126)
Tri3	6 (8/126)
Del7q	27 (34/126)
Gain 12q	10 (12/126)
Tri12	13 (16/126)
Complex genome (≥3 CNAs)	6 (8/126)
SMZL-HR subgroup	53 (67/126)
TL, median (range), kb	3.08 (2.38-7.57)
AH classification	23 (29/126)
EpiCMIT, median (range)	0.745 (0.37-0.91)

Del, Deletion; Tri, Trisomy.

cohort). Mutations were identified using custom pipelines (supplemental Methods). Patients were assigned a mutation-based classification, termed NNK (NF-κB, NOTCH, *KLF2*)-like, based upon the findings of Bonfiglio et al (supplemental Methods).[10]

TL analysis

Telomere length (TL) was quantified in 108 patients using monochrome multiplex polymerase chain reaction (supplemental Methods).[14] Eighty-two chronic lymphocytic leukemia (CLL) patients with single TL analysis data were used for absolute TL extrapolation.[15] Median TL (3.08 kb) across the TL cohort was applied as the cutoff for long vs short TL.

Transcriptomic analysis

Poly(A)-enriched mRNA sequencing libraries were prepared from 15 patients using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, Ipswich, MA) before sequencing on an Illumina NovaSeq 6000. Data were aligned and counted using STAR aligner version 2.7.10b. Downstream analysis was performed in R version 4.3.1 (RRID:SCR_001905) using the edgeR and Gene set enrichment analysis packages (supplemental Methods).[16-19]

DNA methylation profiling

DNA from 126 samples was processed using the Infinium HumanMethylation450 BeadChip (n = 95) or Infinium MethylationEPIC BeadChip (n = 31) (Illumina, Hayward, CA) platforms, as previously reported.[20,21] Data processing was performed using RnBeads version 2.93 (RRID:SCR_010958).[22] Tumor purity was estimated using leukocytes unmethylation for purity[23] and was highly consistent with fluorescence-activated cell sorting–based purity scores ($P < .001$) (supplemental Figure 2). CNAs were identified using conumee.[22] Clusters were identified by hierarchical clustering of M values from the 10 000 most variable CpG sites.[24] This was selected as the optimal approach after comparison with alternative methods (supplemental Methods). EpiCMIT scores were calculated as per Duran-Ferrer et al.[11] patients were assigned to the Arribas-high (AH; n = 29) or Arribas-low (n = 97) methylation subgroup as described by Arribas et al[9] (supplemental Methods). For patients with methylation, CNA, and mutational data available (n = 122), clonal cell fraction (CCF) was calculated for all somatic variants, considering variant allele frequency, local CNA, and purity information.[25,26]

Statistics

Relationships between common clinicobiological variables (defined as ≥2% of patients [n = 5]) were compared using the Fisher exact test, Mann-Whitney U test, or Pearson correlation coefficient (significance, $P = .05$). Univariable Cox regression analysis of differences in time to first treatment (TTFT; months from date of diagnosis to date of first treatment or last follow-up) and OS (months from date of diagnosis to date of death from any cause or last follow-up) was performed (significance; Benjamini-Hochberg–corrected $P = .05$). Kaplan-Meier analysis was used to evaluate OS and TTFT. Multivariate Cox proportional hazards models were generated for TTFT using stepwise backward elimination. Analyses were performed in R version 3.6.1 (RRID:SCR_001905).

Results

WGS identifies disease-specific mutations

An overview of our cohort designs and methodological approaches is shown in Figure 1A. In our cohort of 24 patients

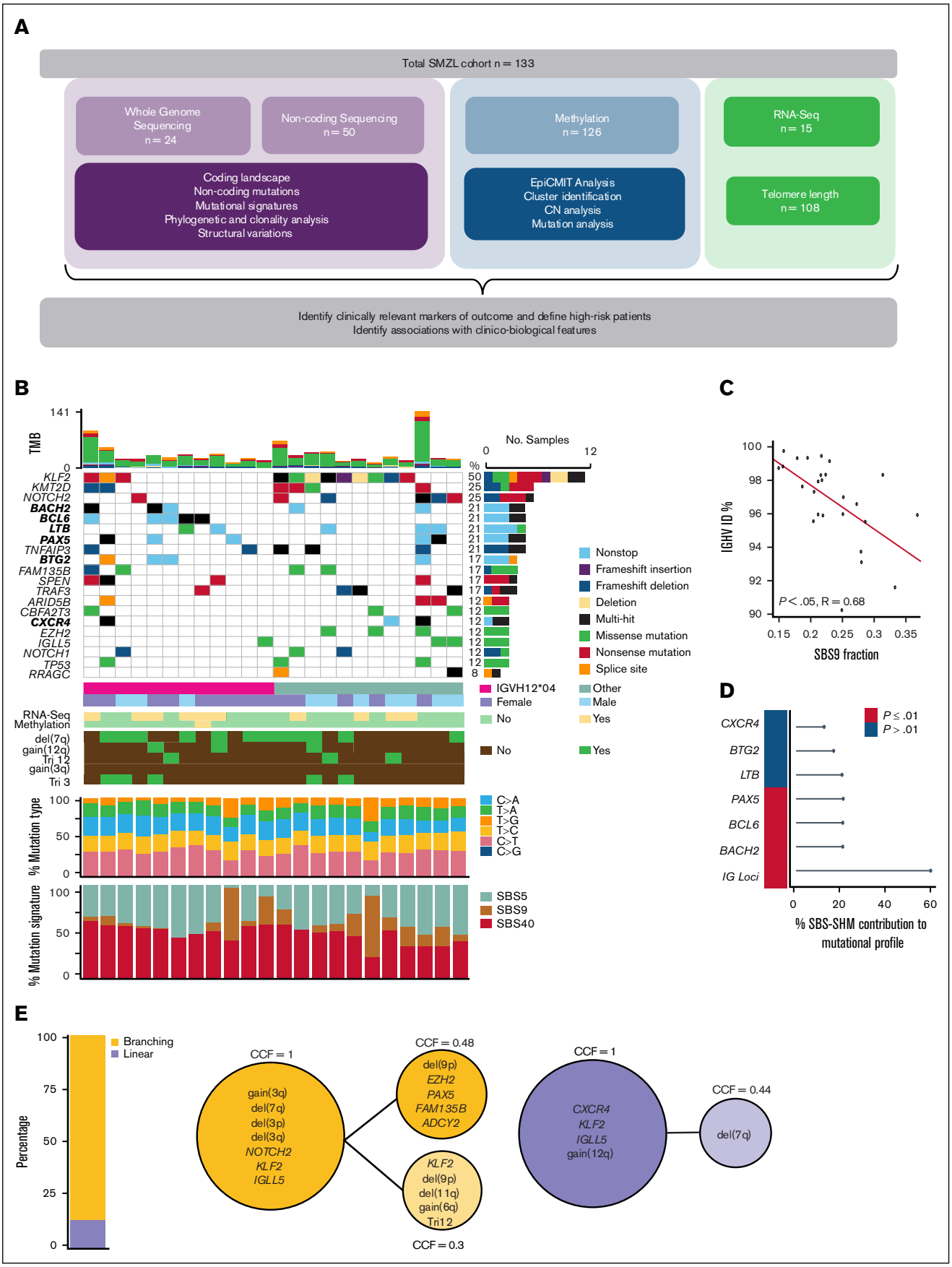
A
Total SMZL cohort n = 133
Whole Genome Sequencing n = 24
Non-coding Sequencing n = 50
Coding landscape
Non-coding mutations
Mutational signatures
Phylogenetic and clonality analysis
Structural variations
Methylation n = 126
EpiCMIT Analysis
Cluster identification
CN analysis
Mutation analysis
RNA-Seq n = 15
Telomere length n = 108
Identify clinically relevant markers of outcome and define high-risk patients
Identify associations with clinico-biological features
B
TMB
No. Samples
KLF2
KMT2D
NOTCH2
BACH2
BCL6
LTB
PAX5
TNFAIP3
BTG2
FAM135B
SPEN
TRAF3
ARID5B
CBFA2T3
CXCR4
EZH2
IGLL5
NOTCH1
TP53
RRAGC
Nonstop
Frameshift insertion
Frameshift deletion
Deletion
Multi-hit
Missense mutation
Nonsense mutation
Splice site
IGVH12*04
Female
Other
Male
RNA-Seq
Methylation
del(7q)
gain(12q)
Tri 12
gain(3q)
Tri 3
No
Yes
% Mutation type
C>A
T>A
T>G
T>C
C>T
C>G
% Mutation signature
SBS5
SBS9
SBS40
C
IGHV ID %
SBS9 fraction
P < .05, R = 0.68
D
CXCR4
BTG2
LTB
PAX5
BCL6
BACH2
IG Loci
P ≤ .01
P > .01
% SBS-SHM contribution to mutational profile
E
Percentage
Branching
Linear
CCF = 1
gain(3q)
del(7q)
del(3p)
del(3q)
NOTCH2
KLF2
IGLL5
CCF = 0.48
del(9p)
EZH2
PAX5
FAM135B
ADCY2
KLF2
del(9p)
del(11q)
gain(6q)
Tri12
CCF = 0.3
CCF = 1
CXCR4
KLF2
IGLL5
gain(12q)
CCF = 0.44
del(7q)

Figure 1.

with matched-normal WGS, we detected a median of 4175 coding somatic mutations per sample (range, 2264-23 219). Consistent with previous studies, the most frequent coding mutations were identified in *KLF2* (50%), *KMT2D* (25%), *NOTCH2* (25%), *TNFAIP3* (21%), and *TRAF3* (17%).[13,27-30] Less frequent recurrent mutations were identified in *EZH2* (12%), *FLNC* (8%), and *BCL10* (4%), together with novel mutations in *FAM135B* (17%), *ARID5B* (12%), and *RRAGC* (8%) (Figure 1B).

Noncoding mutations represent 99.3% (129 559/130 356) of detected variants in our WGS cohort, including recurrent mutations in 5′ untranslated regions (UTRs) of *BACH2* (17%) and 3′ UTRs of *DSG3* (12.5%) and *FGF13* (8%). Mutational hot spots, or kataegis regions, were identified in introns of *BACH2*, *BCL6*, *PAX5*, *BTG2*, *CXCR4*, and *LTB* (Figure 1B). We identified mutational reference signatures from the Catalogue of Somatic Mutations in Cancer in SMZL: single-base substitution (SBS) 40 and SBS5 (both age-related) and SBS9 (linked to polymerase η activity of somatic hypermutation [SHM]) (Figure 1B). The SBS5 signature contributed the most to the mutational landscape (median = 41.6%). A significant correlation was observed between increasing IGHV mutational burden and increased SBS9 signature proportion in individual samples ($P < .05$; $R = 0.68$) (Figure 1C). We also detected the significant enrichment of SHM signature (single base substitution–SHM) described by Machado et al[12] that is associated with activation-induced cytidine deaminase in the germinal center within noncoding mutational hot spots of *BCL6*, *BACH2*, *PAX5*, and *KLF2* coding mutations (Figure 1D). Based on our WGS analysis, a targeted extension panel was designed to validate findings in an additional 50 patients (WGS extension cohort). This revealed at least 1 noncoding hot spot in 69% of patients with recurrent mutations in intron 1 of *BCL6* (26%) and *PAX5* (33%) (supplemental Table 1). In addition, 2 novel missense mutations were found in the *EZH2* coding region, bringing the total mutation count to 5 (p.I55M, p.K199N, p.A255T, p.A692V, and GOF mutation p.Y646F).

Phylogenetic trees were inferred where possible for 21 patients with purity >0.7 and somatic Single Nucleotide Variants (SNV) >3000. Branching subclonal trajectories were most common (86%; n = 18), with linear trajectories observed in only 3 patients (Figure 1E). Notably, del(7q) and gain(3q) CNAs (n = 12 and 5, respectively) originated from the main clonal population in 80% of patients. Similarly, mutations in *KLF2* were always found in the original clone with CCFs of 1.0. In contrast, there was greater variability in clonality of other coding drivers, including *TRAF3* and *KMT2D*, and within our noncoding hot spots.

CNA and structural variations in SMZL

After CNA calling, using both methylation array (n = 126) and paired WGS data, we revealed at least 1 CNA (range, 0-14) in 73% of patients, with 29 patients (23%) harboring a complex genome (≥3 CNAs per genome). The most prevalent CNAs were del(7q) (27%), gain(3q) (16.6%), trisomy 12 (12.7%), and gain(12q) (9.5%) (Figure 2A). Minimally deleted or gained regions were identified on chromosomes 3q, 6q, 8p, 12q, 13q, and 17p (supplemental Table 2). All patients with deletions of chromosome 17p (n = 8) had concomitant *TP53* mutations. Although chromosome 7q deletion breakpoints were heterogeneous in both size and location, a 1.75-Mb minimally deleted region (128.57-130.32 Mb) was identified, which encompassed 24 coding genes, including *FLNC*, and 12 noncoding RNAs (Figure 2B). *FLNC* mutations were all missense (n = 11; 4%), with a single patient harboring biallelic loss with both a deletion and a missense mutation (p.P1060L) (Figure 2C).

Using our WGS patients (n = 24), we focused on the further characterization of 33 SVs localized to regions of recurrent CNAs and the immunoglobulin heavy chain locus. Twenty translocations targeted chromosomes 3 (n = 8), 7 (n = 5), 12 (n = 5), and 14 (n = 2) (supplemental Table 3). Of the 12 WGS patients with del(7q), 4 were shown to be the result of a derivative chromosome. In patient T32, del(7q) was the result of der(7)t(7,15)(q32.1;15q.1) with breakpoints in intron 1 of *RTF1* (chr15:41425102) and exon 2 of *ATP6V1FNB* (chr7:128870589). In patient T14, we identified der(7)t(7,12)(q31.1;q13.12) with a breakpoint in intron 7 of *NRCAM* (chr7:108 227 210) and duplication of 12q (chr12:51040961). Finally, in patient T43, deletion of 7q and duplication of 3q were the result of der(7)t(3,7)(q13.13;q31.31) with breakpoints in intron 1 of *KCND2* (chr7:120299899) and intron 14 of *MORC1* (chr3:109040204) (Figure 2D). Interestingly, in patient T7T2, we identified a der(9)t(9;14)(p13.3;q32.33), which results in translocation between *PAX5* intron 1 (chr9:37033006) and the immunoglobulin heavy chain locus (chr14:105709625), a novel observation in SMZL. Inversions were less common (n = 13), and notably, 2 patients harbored a 6.9-Mb chromosome 7 inversion with breakpoints occurring within 30 base pairs of each other (patients T7T2 and T32) (supplemental Table 3). In both patients, inversion resulted in disruption of the lncRNA LOC105375294 and the *ZYX* genes.

Finally, we calculated the CCF for somatic variants in patients with CNA data (n = 122). Mutations within coding drivers *KLF2*, *KMT2D*, and *NOTCH2* were predominantly clonal, with average CCF greater than 0.8. In contrast, *SPEN*, *TRAF3*, *ARID5B*, and *TP53* had primarily subclonal mutations, though a wide distribution of CCFs was observed across most genes (Figure 2E). In 5 of 6 samples with multiple *KLF2* mutations, at least 1 mutation was clonal (CCF > 0.8), with additional subclonal mutations likely representing later events, often with lower CCFs. Clonal mutations were predominantly missense (n = 4/6), whereas subclonal mutations were typically frameshift (n = 4/6) (Figure 2F).

Figure 1. SMZL WGS, noncoding landscape, mutational signatures, and subclonality. (A) CONSORT diagram showing the experimental workflow and key analytical processes applied in our study. (B) Waterfall plot showing the top 20 most frequently mutated genes/regions in the SMZL WGS cohort (n = 24), with noncoding mutations in bold. The horizontal bars below indicate IGHV use, sex, matched analyses, and cytogenetic features for each patient. Bar charts show the base-pair context for mutations and the contribution of SBS signatures to the mutational landscape. (C) Scatterplot showing the significant negative Pearson correlation ($P < .05$) between IGHV mutational burden and the SBS9 signature fraction in the mutational spectra of individual SMZL samples. The linear regression model is shown by the red line. (D) Bar chart showing significant enrichment (red) of the SBS-SHM signature in noncoding mutations compared with the genome-wide mutational background. (E) Bar graph showing the percentage of patients studied with a branching (yellow) or linear (lilac) subclonal trajectory, with genomic profiles of an example of each trajectory type. CCF indicates cancer cell fraction of the subclonal tumor population or node. del, deletion; tri, trisomy.

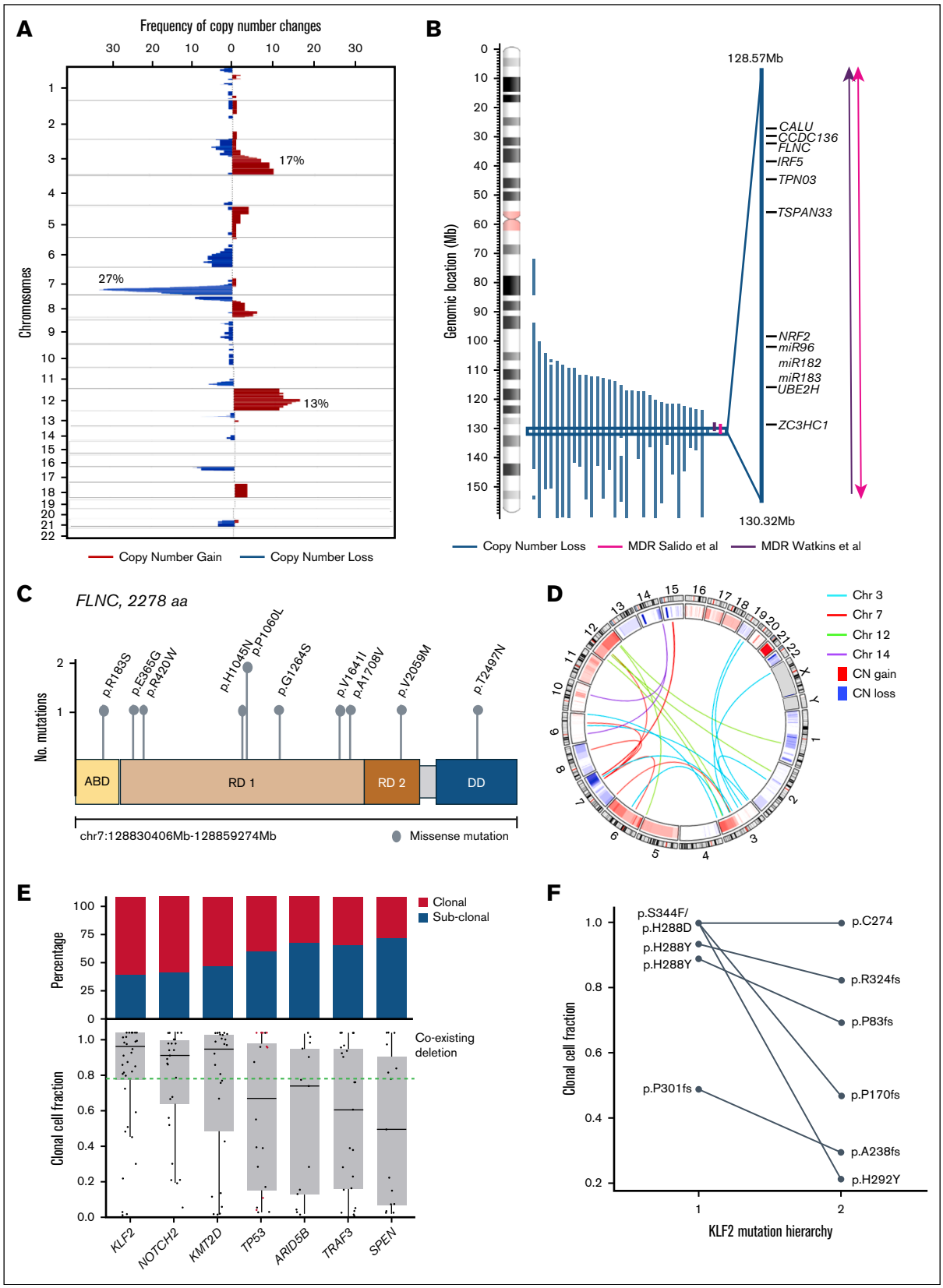

Figure 2. Coding landscape in the integrated cohort. (A) Frequency plot showing genome-wide copy number gains and losses. (B) Ideogram of chromosome 7q deletions. Blue bars represent genomic location and size of individual deletions. The blue box indicates the minimally deleted region, and key genes located within this region are provided

Presence of 2 epigenetically defined patient subgroups with differing clinicobiological features

Next, we aimed to define epigenetically distinct subtypes of SMZL using methylation array data. In the absence of a validation cohort, the methylation cohort was randomly divided into discovery (n = 63) and validation (n = 63) sets. Hierarchical clustering of the 10 000 most variable CpG sites identified 2 clusters in both groups. Bootstrapped survival analyses confirmed TTFT differences (supplemental Methods; supplemental Figure 3). Subsequently, we performed hierarchical clustering on the entire cohort: high-risk (HR) SMZL (n = 67; 53% patients) and low-risk (LR) SMZL (n = 59; 47% patients). We identified 6891 differentially methylated CpG sites, including 6352 hypermethylated and 541 hypomethylated CpGs in the SMZL-HR group (Figure 3A-B).

Sixteen clinicobiological features were significantly enriched in SMZL-HR patients, including female sex ($P < .01$), short TL ($P < .001$), key driver gene mutations (*KLF2* [$P < .001$], *KMT2D* [$P < .0001$], *TRAF3* [$P = .004$], *NOTCH2* [$P = .02$], *BCL10* [$P = .006$], *FLNC* [$P = .02$]) (Figure 3C) and chromosomal alterations (del[7q], gain[3q], gain[12q] [all CNAs, $P < .01$]) (Figure 3D). Crucially, all patients with IGHV1-2*04 use were assigned to the SMZL-HR subgroup ($P < .001$). The fraction of SBS40 mutational signatures was also increased in SMZL-HR compared with SMZL-LR ($P < .001$) (supplemental Figure 5). In contrast, SMZL-LR was significantly associated with high IGHV mutational load (<97% germ line identity; $P < .001$; Figure 3C) and showed a nonsignificant trend toward trisomy 12 ($P = .06$). Distribution of blood and spleen samples was similar between epitypes ($P = .34$), suggesting that observed biological differences are unlikely to be driven by tissue of origin.

SMZL-HR epitype exhibits evidence of elevated cell proliferation

We next calculated the epiCMIT proliferative history score[11] for each patient (Figure 4A). We observed that patients with SMZL-HR exhibited significantly higher epiCMIT values compared with patients with SMZL-LR ($P < .001$; Figure 4B). They were also significantly linked to the presence of IGHV1-2*04 gene use ($P = .007$; Figure 4B) and positively correlated with TMB ($r = 0.35$; $P < .001$; Figure 4C). TL data were available on 101 patients and ranged from 2.38 to 7.57 kb (median, 3.06 kb; Figure 4Di). Interestingly, we observed a significant negative correlation between epiCMIT and TL ($R = -0.3$; $P = .002$), supporting an association between cell proliferation and concomitant telomere attrition (Figure 4Dii). Transcriptomic comparisons between SMZL-HR (n = 10) and SMZL-LR (n = 5) identified 399 differentially expressed genes (232 underexpressed and 167 overexpressed; False Discovery Rate (FDR) < .05; log fold change ≥1.5). Gene set enrichment analysis identified enrichment of 2 pathways associated with elevated cell division; E2F targets (normalized enrichment score [NES], 1.98; $P < .01$) and G2M checkpoint (NES, 2.07; $P < .01$), as well as epigenetic regulation through KAMMINGA_*EZH2* targets (NES, 1.91; $P = .006$) (supplemental Table 4).

Finally, we included epiCMIT in a comprehensive correlation analysis of 63 variables (supplemental Table 5) with significant associations shown in Figure 4E and supplemental Table 6. A total of 18 co-occurring associations (FDR < .05) were detected, many confirming the significant associations of SMZL-HR with previously discussed clinicobiological variables. In addition, significant associations were observed between *KLF2* mutations, del(7q) and IGHV1-2*04, and *BCL10 and TNFAIP3* mutations. Co-occurrences were noted for trisomies of chromosomes 3 and 18. IGHV mutational loads <97% were mutually exclusive from IGHV1-2*04 use, SMZL-HR status, and *KLF2* mutations.

Clinical implications of molecular disease features, including the SMZL-HR subgroup

During the follow-up period (median, 5 years; range, 0-24), 95 of 133 patients required treatment (71%). Among these, 41% received rituximab (with or without chemotherapy), and 42% underwent splenectomy as first-line therapy. A subset of 21% of patients required subsequent treatment, with rituximab being the preferred therapy in 65% of patients. Thirty-six patients died during the study. Although data on disease transformation were limited (n = 17), we observed associations with high-risk features, including IGHV1-2*04 use ($P = .005$), del(7q) ($P = .003$), SMZL-HR ($P = .009$), *KLF2* ($P = .02$), and mutation of *TP53* ($P = .03$).

Univariable Cox regression analysis of 55 clinical and molecular features was performed to evaluate prognostic significance for TTFT and OS (supplemental Tables 7 and 8), with significant associations displayed in Figure 5A. Key findings showed that SMZL-HR status (hazard ratio, 1.9; $P = .003$), IGHV1-2*04 (hazard ration, 2.1; $P = .003$), *NOTCH2* mutations (hazard ratio, 1.9; $P = .013$), female sex (hazard ratio, 1.8; $P = .005$), and gain of 3q (hazard ratio, 1.8; $P = .003$) were significantly associated with shorter TTFT. *NOTCH1* mutations were included in a panel of 7 genes (*CARD11*, *KLF2*, *NFKBIE*, *NOTCH1*, *NOTCH2*, *TNFAIP3*, and *SPEN*) used by Bonfiglio et al to assign patients to a high-risk subgroup they termed NNK.[10] Our patients were similarly defined as NNK-like (n = 78) or "other" (n = 51) based on the mutation status of these genes. In addition, to compare our model to the 27K promoter array classification defined by Arribas in 2015, we assigned patients an AH methylation (n = 29) and Arribas-low methylation (n = 97) status using 86 of the top 100 most variable CpGs identified by the authors[11] (supplemental Table 9), showing that the AH group was significantly associated with shorter TTFT (hazard ratio, 1.8; $P = .02$). Significant predictors of shorter OS include the SMZL-HR epitype (hazard ratio, 2.5; $P = .039$), age >60 years (hazard ratio, 18.7; $P < .001$), del(7q) (hazard ratio, 2.3; $P = .02$), and genomic complexity (>3 CNA) (hazard ratio, 2.8; $P = .0038$).

Figure 2 (continued) to the right. The pink and purple lines indicate previously described minimally deleted regions (127.5-130.8 Mb[31] and 127.03-130.07 Mb,[32] respectively). (C) *FLNC* lollipop plot showing the locations of 10 missense mutations (gray circles). (D) Circos plot displaying SVs originating from chromosomes 3 (blue lines), 7 (red lines), 12 (green lines), and 14 (purple lines). Chromosome ideograms are displayed in the outer panel, with copy number changes (blue, loss; gain, red) shown in the inner panel. (E) Clonality of coding mutations in 7 driver genes, displayed in order of decreasing clonality. Bar charts (top panel) show the average percentage of cells in mutated patients that are clonal (red) vs subclonal (blue), whereas the box-and-whisker plots below show distribution of cancer cell fractions. Mutations with CCF >0.8, as indicated by the dotted green line, are considered clonal. *TP53* mutations that coexist with deletion of the second allele are indicated in red. (F) Trajectory plot of CCFs in double-hit *KLF2* mutants. ABD, actin binding domain; Chr, chromosome; CN, copy number; DD, dimerization domain; RD, rod domain.

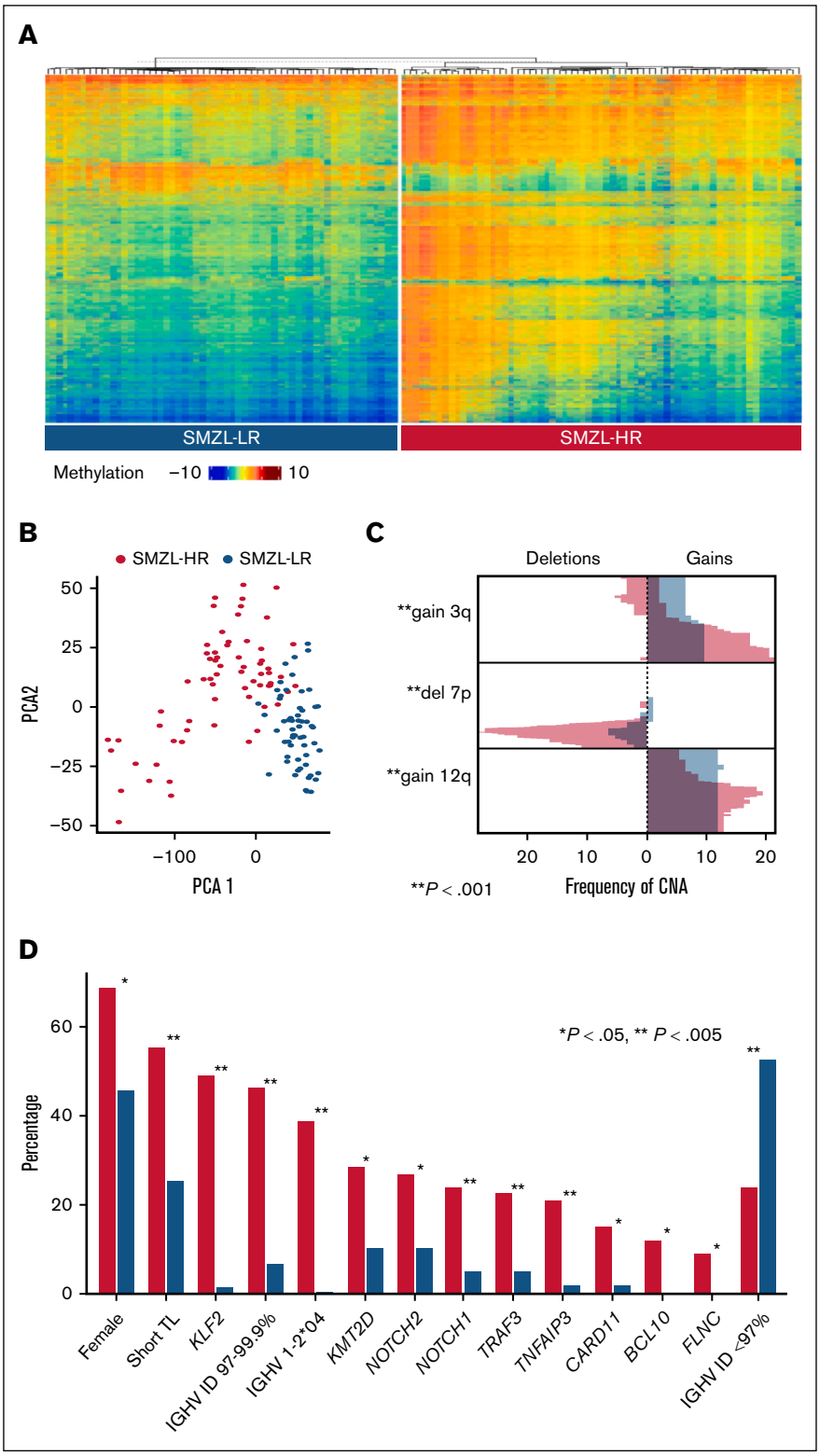

Figure 3. Clustering and association with clinicobiological features. (A) Heat map, showing the top 10 000 most variable CpG sites that were used in clustering of SMZL samples, with each patient represented on the x-axis. Samples and CpGs are clustered using a complete Euclidean distance method with hierarchical clustering. The horizontal bars below show the clusters (SMZL1 [HR], burgundy; SMZL2 [LR], blue). (B) Scatterplot displaying results from principal component analysis of the top 10 000 most variable CpG sites used for clustering. (C) Frequency plot showing copy number gains and deletions (right- and left-hand panel, respectively) significantly associated with patients with SMZL-HR and SMZL-LR. (D) Frequency bar chart showing features that are significantly associated with patients with SMZL-HR (burgundy) and SMZL-LR (blue). Significant enrichment of all features was determined using Fisher exact or χ^2 test, with significance levels indicated by $*P > .05$; $**P < .005$. PCA, principal component analysis.

Kaplan-Meier analysis further supported the prognostic value of the SMZL-HR entity, showing significant differences in both TTFT ($P = .0034$) and OS ($P = .039$) between the SMZL-HR and SMZL-LR groups (Figure 5B). Patients with SMZL-HR also exhibited a higher frequency of therapeutic intervention ($P < .001$), disease transformation ($P = .007$), and increased mortality ($P < .001$) than patients with SMZL-LR (Figure 5C). In analyses restricted to patients diagnosed since 2010, significant

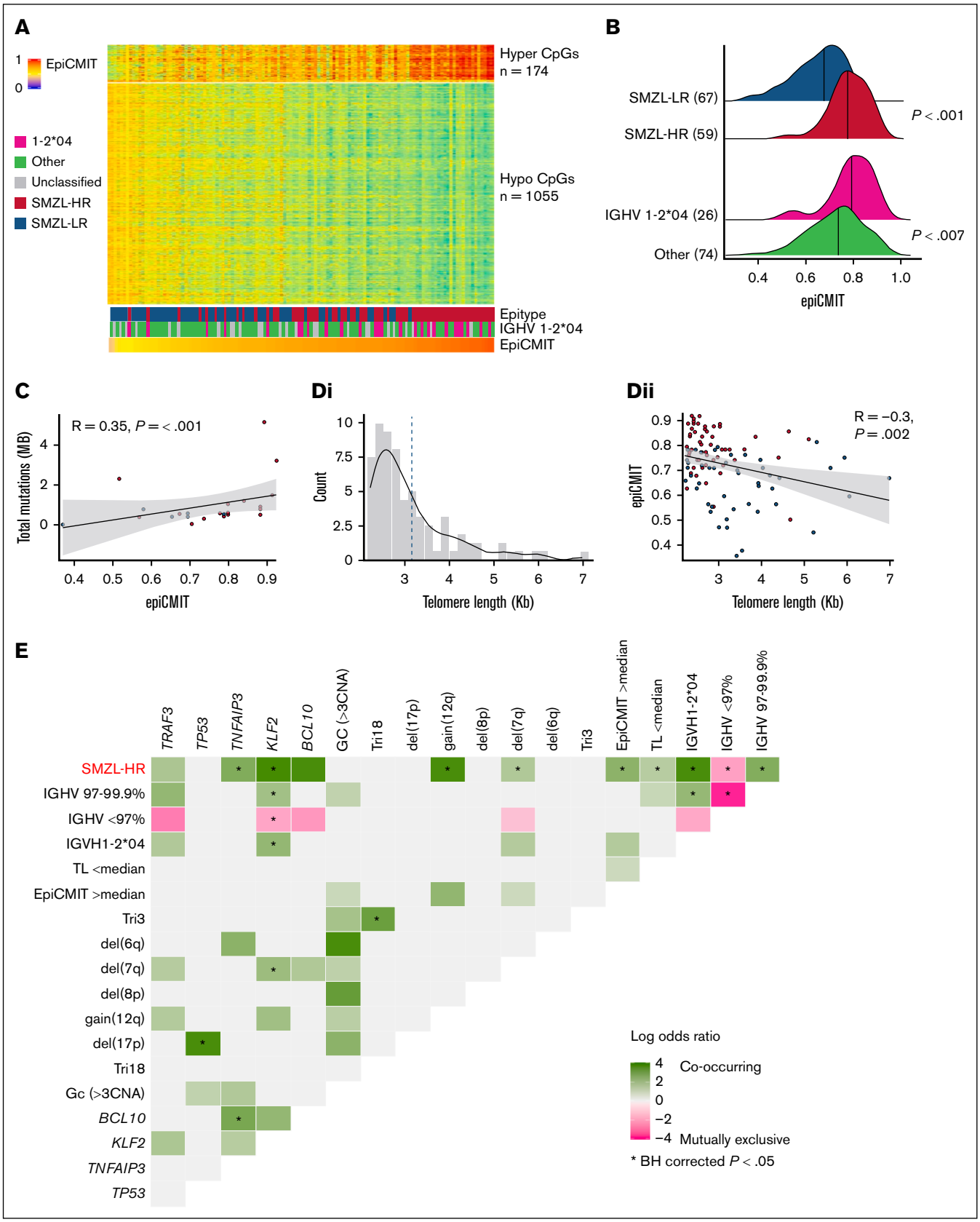

Figure 4. EpiCMIT and association with clinicobiological features. (A) Heat map showing methylation status of 174 and 1055 CpGs used to calculate epiCMIT-hyper and epiCMIT-hypo scores, with each patient represented on the x-axis. The horizontal bars below show EpiCMIT score, cluster, and IGHV1-2*04 status for each patient. (B) Ridge

differences in TTFT and OS between epitypes were maintained, supporting the continued prognostic relevance of the score in the modern treatment era (supplemental Figure 5). SMZL-HR identified more poor-risk patients (n = 63) than the Arribas classifier (n = 29)[9]; Uniform Manifold Approximation and Projection (UMAP) analysis demonstrated that the Arribas classifier was able to capture only the most distinctly methylated cases from the SMZL-HR group (Figure 5D), with more subtle methylation signatures missed by CpGs selected using the deprecated 27K array platform. Kaplan-Meier analysis confirmed that patients with SMZL-HR experienced significantly shorter TTFT than patients with SMZL-LR regardless of their Arribas classification (Figure 5E). These findings represent a novel advancement of existing classifiers, where SMZL-HR identifies patients with high-risk missed by previous classifiers and is relevant for modern methylation assay technologies.

We estimated the adjusted impact of the SMZL-HR epitype on TTFT using multivariate Cox proportional hazard analysis. The model incorporated the 6 variables predictive of shorter TTFT, as determined by univariate analysis (SMZL-HR, IGHV1-2*04, female sex, AH, *NOTCH2* mutation, and gain of 3q). It was built using data from 91 patients with complete data, encompassing a total of 65 events. In the final model, SMZL-HR (hazard ratio, 2.4; $P = .001$) and *NOTCH2* mutations (hazard ratio, 1.84; $P = .04$) remained the only significant independent variables associated with TTFT. We also developed a model to assess the SMZL-HR subgroup in the context of the Arribas classification, epiCMIT scores, and the NNK-like classifier. SMZL-HR remained the sole significant independent variable in the final model (hazard ratio, 1.96; $P = .000$), which was based on 115 patients with 84 events. Because of the low number of events, multivariate Cox proportional hazard analysis for OS was not conducted.

Discussion

Here, we analyzed the (epi)genomic landscape of SMZL, providing novel insights into its molecular and clinical features. Our WGS study, which to our knowledge is the first of its kind to include matched-normal duos, provides crucial insights into the coding and noncoding mutational landscape of this lymphoma. We identified a median of 1.4 mutations per megabase of genomic DNA, higher than in previous SMZL exome studies[33] and other indolent B-cell neoplasms, such as mantle cell lymphoma,[34] CLL,[35] and hairy cell leukemia,[36] and more similar to certain diffuse large B-cell lymphoma[37] and multiple myeloma[38] studies. In addition to coding mutations in known driver genes, we identified novel recurrent mutations in *FAM135B* and *RRAGC*. Although *RRAGC* mutations have been reported in 9% to 18% of follicular lymphoma,[39,40] where they have a role in deregulating mechanistic target of rapamycin signaling downstream of nutrient sensing and induce murine lymphomagenesis,[41,42] we found the same *RRAGC* variants in SMZL (n = 3, 8%).[39,43] Notably, 4% of patients harbor missense mutations in FLNC, confirming it as a target of recurrent mutations in SMZL and highlighting the need for future mechanistic studies.[13,27] *FLNC* has been linked to various cancers, including gastric and prostate cancer, where methylation and transcriptomic changes relate to tumor progression.[44,45,46-49] We also characterized recurrent mutations in the *EZH2* gene, including Y646F and A255T, previously observed in other hematological malignancies.[50] Y646F mutation is a known biomarker for tazemetostat responsiveness and transformation in patients with follicular lymphoma.[51,52] However, K199T mutation in the PRC2 domain is newly identified in SMZL and has not been reported in other B-cell malignancies, marking a novel discovery. Our phylogenetic analysis revealed predominantly branching subclonal trajectories in SMZL. Although *KLF2* mutations typically arise from the main clonal population, several patients exhibited convergent *KLF2* lesions, highlighted by 1 clonal and 1 subclonal mutation, underscoring the critical role of this gene in SMZL lymphomagenesis.

Recurrent noncoding mutations were identified in 5′ UTRs of *BACH2* and *BCL6* and 3′ UTRs of *PAX5*, *DSG3*, and *FGF13*. Although these mutations are previously unreported in SMZL, similar patterns have been reported in related B-cell malignancies, such as *BACH2* promoter mutations in CLL,[25] *PAX5*,[53] and *BCL6* noncoding mutations in various B-cell lymphomas.[54-56] Validation in an additional cohort of 50 patients confirmed noncoding hot spot mutations in 69% of patients, underscoring their relevance in the disease. We identified 3 mutational signatures linked to aberrant SHM, with a significant correlation between IGHV mutational burden and the SBS9 signature. Notably, we observed a novel enrichment of the SHM signature in *KLF2* coding mutations, suggesting that aberrant SHM may play a role in the acquisition of these mutations. Although this is the first report in SMZL, aberrant SHM is well-documented in other lymphomas,[57-61] where activation-induced cytidine deaminase misfires introduce mutations in key genes for B-cell function and transformation.[61]

Hierarchical clustering of methylation data and survival analysis revealed 2 distinct patient subgroups: SMZL-HR and SMZL-LR, with patients with SMZL-HR experiencing markedly shorter TTFT. This high-risk subgroup exhibited adverse features, including recurrent mutations in key driver genes, such as *KLF2*, *NOTCH2*, *KMT2D*, and *TRAF3*, along with chromosomal alterations like del(7q), gain(3q), and gain(12q). Notably, although SMZL-HR was associated with structural alterations affecting the 12q chromosome, the SMZL-LR subgroup demonstrated complete trisomy 12, suggesting potentially convergent evolution targeting this chromosome. Furthermore, the SMZL-HR epitype demonstrated a heightened history of, and potential for, cell proliferation, as

Figure 4 (continued) plots showing distribution of epiCMIT scores for variables exhibiting statistically significant differences between 2 status groups. The number of patients with each variable is shown. *$P < .05$; **$P < .005$. (C) Scatterplot showing significant positive Pearson correlation ($P < .001$) between epiCMIT score and TMB per megabase in coding regions. The linear regression model is depicted by the black line, and the gray area indicates confidence intervals. (Di) Histogram showing frequency and distribution of TL (kb) in our cohort, with median length of 3.1 kb represented by the blue dotted line. (Dii) Scatterplot showing negative Pearson correlation ($P = .001$) between high epiCMIT score and TL. Points are colored to indicate SMZL epitype (burgundy, SMZL-HR; blue, SMZL-LR). (E) Association plot of log odds ratios showing SMZL features that co-occur or are mutually exclusive (green and pink squares, respectively); corrected *P* values calculated using the Benjamini-Hochberg (BH) false discovery rate, $P \leq .05$). del, deletion; tri, trisomy.

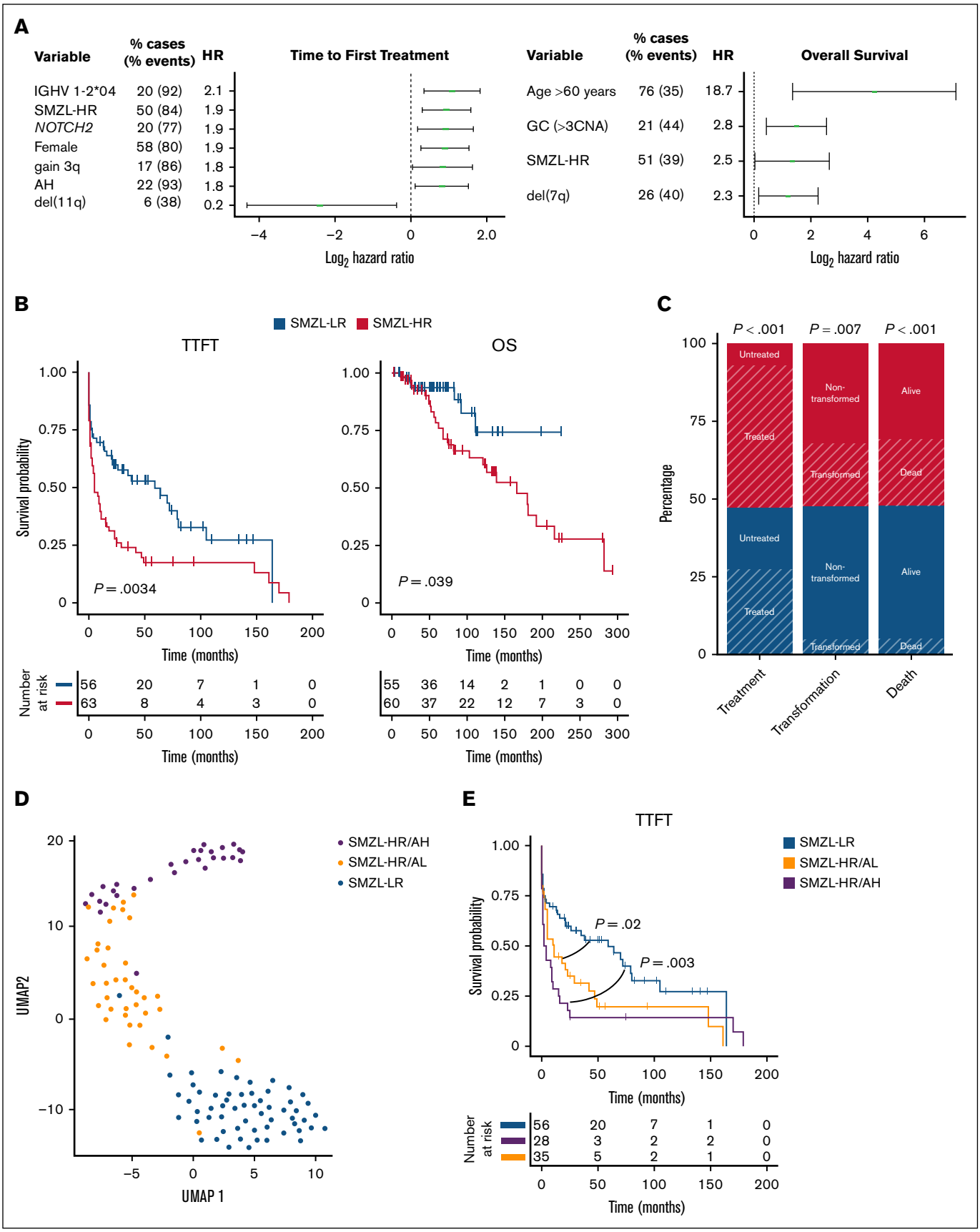
A
Variable
% cases (% events)
HR
Time to First Treatment
IGHV 1-2*04 20 (92) 2.1
SMZL-HR 50 (84) 1.9
NOTCH2 20 (77) 1.9
Female 58 (80) 1.9
gain 3q 17 (86) 1.8
AH 22 (93) 1.8
del(11q) 6 (38) 0.2
−4 −2 0 2.0
Log2 hazard ratio
Variable
% cases (% events)
HR
Overall Survival
Age >60 years 76 (35) 18.7
GC (>3CNA) 21 (44) 2.8
SMZL-HR 51 (39) 2.5
del(7q) 26 (40) 2.3
0 2 4 6
Log2 hazard ratio
B
SMZL-LR
SMZL-HR
TTFT
OS
Survival probability
1.00
0.75
0.50
0.25
0
P = .0034
P = .039
0 50 100 150 200
0 50 100 150 200 250 300
Time (months)
Number at risk
56 20 7 1 0
63 8 4 3 0
55 36 14 2 1 0 0
60 37 22 12 7 3 0
C
P < .001
P = .007
P < .001
Percentage
100
75
50
25
0
Untreated
Treated
Non-transformed
Transformed
Alive
Dead
Treatment
Transformation
Death
D
SMZL-HR/AH
SMZL-HR/AL
SMZL-LR
UMAP2
20
10
0
−10
−5 0 5 10
UMAP 1
E
TTFT
SMZL-LR
SMZL-HR/AL
SMZL-HR/AH
Survival probability
1.00
0.75
0.50
0.25
0
P = .02
P = .003
0 50 100 150 200
Time (months)
Number at risk
56 20 7 1 0
28 3 2 2 0
35 5 2 1 0
0 50 100 150 200
Time (months)

Figure 5.

evidenced by elevated epiCMIT scores. We also found a positive correlation between TMB and epiCMIT scores, underscoring the relationship between cellular proliferation and accumulation of somatic mutations. Conversely, a negative correlation between TL and epiCMIT scores suggests that increased cellular proliferation may coincide with telomere attrition, potentially driving clonal evolution and disease progression. To our knowledge, this study marks the first investigation of TL and epiCMIT in SMZL, revealing that epiCMIT serves as a more robust biomarker than TL, which can be influenced by transient telomerase upregulation in germinal center-derived B cells.[62] In addition, transcriptomic analysis identified differentially expressed genes within the SMZL-HR subgroup involved in cell division pathways, further supporting the concept that patients with SMZL-HR also possess the potential for future cellular division that likely drives the progressive phenotype observed in these patients.

By identifying 2 distinct disease entities, we revealed significant differences in their mutational, clinicobiological, and transcriptomic profiles that affect clinical behavior. Although the SMZL-HR subtype emerged as an independent predictor of TTFT, it did not reach significance for OS, likely because of treatment heterogeneity and unrelated causes of death in elderly patients. Our SMZL-HR epitype substantially extends the important observations of Arribas et al[9] by delivering a more comprehensive DNA-methylation–based definition of the high-risk subgroup. This refined SMZL-HR category identifies a markedly larger cohort of poor-risk patients and offers a more granular view of the underlying biology encompassing coding and noncoding mutational profiles, mutational processes, co-occurrence with recurrent genomic lesions, proliferative history, and telomere dynamics. Importantly, the SMZL-HR classifier shows superior discriminatory power in multivariate TTFT models that incorporate previously proposed risk predictors,[9,10,63] highlighting its potential strength as an independent prognostic tool. Although our reconstruction of the Arribas classifier was necessarily limited to 86 CpGs present on modern arrays and may not fully recapitulate their original subgroup, our SMZL-HR classification reliably identifies the same high-promoter-methylation cases. Furthermore, the genome-wide nature of the SMZL-HR epitype extends beyond this promoter-restricted signal, capturing additional patients with more subtle methylation differences but comparably adverse clinicobiological features. In addition, we found a novel association between the SMZL-HR epitype and elevated epiCMIT levels, female sex, and IGHV1-2*04 use. Recent findings suggest that SMZL patients expressing the IGHV1-2*04 gene exhibit a distinct immunogenetic profile influenced by microenvironmental factors.[64] Since our SMZL-HR epitype included all IGHV1-0*24 patients, it may represent a biologically distinct subgroup of SMZL with unique epigenetic and immunogenetic characteristics. However, thorough immunogenetic profiling of non-IGHV1-2*04 patients within this epitype is needed to confirm this distinction at the immunogenetic level.

A key strength of our study lies in the integrated analysis of a substantial cohort of patients with this rare lymphoma, incorporating extensive clinical, biological, and immunogenetic data. Although central pathology review was not feasible, diagnostic accuracy was maintained by including only high-confidence cases provided by international experts using established diagnostic criteria. Unlike previous studies that predominantly examined splenic tissue,[9,10] our tumor material was mainly blood-derived and therefore reflects complementary pretreatment disease biology using samples that are more accessible in the modern era, when splenectomy is rare for diagnostic but also therapeutic indications. Although our subgroups should be viewed as translational rather than clinical, because robust external validation in independent cohorts is required, our study significantly advances previous genomic and epigenetic analyses of SMZL[9,10] in 2 principal ways. First, by combining the first tumor-normal WGS data set in SMZL with genome-wide methylation profiling and broader multiomic integration, we identified fundamental biological drivers of the disease, including noncoding aberrant SHM hot spots, detailed clonal architectures, telomere and proliferative dynamics, and coordinated immunogenetic features, none of which were captured in earlier, more limited studies. Second, the resulting SMZL-HR subgroup represents a biologically coherent and clinically meaningful high-risk category characterized by markedly shorter TTFT, increased transformation risk, and enrichment of key adverse genomic lesions, providing a prognostic framework with substantially improved precision and independent predictive value. In summary, our study refines the understanding of SMZL pathogenesis and has direct implications for improving risk stratification and guiding more personalized clinical management strategies in SMZL.

Acknowledgments

The authors gratefully acknowledge the patients studied herein for providing primary tumor and germ line material and clinical data. They also acknowledge the IRIDIS High Performance Computing Facility and associated support services at the University of Southampton, as well as the Faculty of Medicine Bioinformatics Research Facility at the University of Southampton, for assistance with this study. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure Sweden and Science for Life Laboratory.

This study was supported by research grants from Cancer Research UK (ECRIN-M3 accelerator award C42023/A29370), Kay Kendall Leukemia Fund (873, 1104), Blood Cancer UK (11052, 12036), Wessex Medical Research, Bournemouth Leukemia Fund, and AIRC under 5 per Mille 2018 (ID 21198; Principal Investigator Roberto Foà and Group leader Paolo Ghia). A.M. was supported by the Cancer Immunology Talent Fund and the University of Southampton. R.R. was supported by grants from the

Figure 5. Clinical outcome in SMZL is associated with genomic and epigenetic features. (A) Forest plots showing hazard ratios for variables significant ($P < .05$) for TTFT and OS in univariable Cox regression analysis. The proportion of patients positive for each variable and the proportion of events in those patients are shown. (B) Kaplan-Meier curves for TTFT and OS (months) for patients with SMZL-HR and SMZL-LR (burgundy and blue lines, respectively). (C) Stacked bar graph showing the significant difference in proportions of patients with SMZL-HR and SMZL-LR who received treatment, transformed, or died. Mann-Whitney U test with Benjamini-Hochberg correction was used to determine statistical significance ($P < .05$). (D) UMAP showing the distribution of patients classified as SMZL-HR and Arribas High-M (purple), SMZL-HR but Arribas Low-M (orange), and SMZL-LR and Low-M (blue). (E) Kaplan-Meier curve for TTFT for patients with SMZL-HR dichotomized by AH methylation (AH) and Arribas low methylation (AL) (purple and orange lines, respectively), and SMZL-LR (blue line). UMAP, Uniform Manifold Approximation and Projection.

Swedish Cancer Society, the Swedish Research Council, Region Stockholm, and Radiumhemmets Forskningsfonder, Stockholm.

Authorship

Contribution: H.P., D.B., D.O., and J.C.S. designed the study; H.P., A.M., C.J.O., M.J.J.R.Z., H.E.A., L.C., S.S., M.S., M.P., Z.D., N.M.-B., and D.W. performed laboratory work; H.P., D.B., A.M., L.B., J.T., C.J.O., V.L., C.O., and B.S. analyzed data; H.P., D.B., B.S., A.M., D.O., and J.S. wrote the manuscript; D.O., F.F., R.W., K.S., A.X., L.S., P.G., C.T., C.K., G.P., E.K., R.R., G.P., D.R., R.-M.A., S.W., M.A., A.F., and M.S. provided study material and clinical data; D.O., R.W., J.G., L.H., R.M., C.P., S.E., C.O., M.D.-F., and J.I.M.-S. provided expert advice; and all authors reviewed the manuscript.

Conflict-of-interest disclosure: R.R. reports honoraria from AbbVie, AstraZeneca, Illumina, Janssen, and Roche. L.S. reports consultancy for AbbVie, AstraZeneca, BeiGene, Janssen, and Lilly. P.G. reports honoraria, consultancy fees, and research funding from AbbVie, AstraZeneca, BeiGene, Bristol Myers Squibb, Lilly, Janssen, MSD, and Roche. D.R. reports honoraria and research funding from AbbVie, AstraZeneca, BeiGene, Bristol Myers Squibb, Gilead, Janssen, Lilly, and Kite. M.A. reports research funding from Pfizer. F.F. reports honoraria from AbbVie, AstraZeneca, BeiGene, and Janssen. R.W. reports honoraria from AbbVie, AstraZeneca, BeiGene, Janssen, and Secura Bio. C.T. reports honoraria and/or consultancy fees from Bristol Myers Squibb, Janssen, Novartis, Roche, AbbVie, Gilead, Hospira, Amgen, Cellectis, Kite, Takeda, Incyte, and Bayer. J.C.S. reports research funding from Roche Holding AG and Cambridge Epigenetix. The remaining authors declare no competing financial interests.

ORCID profiles: A.M., 0000-0002-3454-8723; B.S., 0009-0001-4516-8945; C.J.O., 0000-0002-9875-0998; M.D.-F., 0000-0003-1666-5819; L.B., 0000-0002-7989-9404; H.E.A., 0009-0003-4012-4577; J.T., 0000-0002-3569-9101; M.J.J.R.-Z., 0000-0002-1064-5350; L.S., 0000-0002-0844-0989; P.G., 0000-0003-3750-7342; G.G., 0000-0002-4681-0151; S.W., 0000-0002-8914-0370; M.A., 0000-0002-2037-3926; M. Seifert, 0000-0001-9453-369X; E.K., 0000-0003-3078-6131; R.R., 0000-0002-0211-8788; R.-M.A., 0000-0003-0901-5252; G.P., 0000-0002-6937-2852; F.F., 0000-0002-2211-1831; M. Salido, 0000-0001-9988-5977; A.F., 0000-0002-3381-9472; L.H., 0000-0002-6413-6586; R.M., 0000-0003-2932-7800; R.W., 0009-0007-5693-2789; J.I.M.-S., 0000-0001-8809-5195; C.O., 0000-0001-8000-9694; D.B., 0000-0003-3163-608X; J.C.S., 0000-0002-0972-2881.

Correspondence: Jonathan C. Strefford, Cancer Genomics Group, Southampton General Hospital, MP824 Somers Building, Tremona Rd, Southampton SO16 6YD, United Kingdom; email: jcs@soton.ac.uk.

References

1. Matutes E, Oscier D, Montalban C, et al. Splenic marginal zone lymphoma proposals for a revision of diagnostic, staging and therapeutic criteria. *Leukemia*. 2008;22(3):487-495.
2. Zucca E, Arcaini L, Buske C, et al; ESMO Guidelines Committee. Electronic address: clinicalguidelines@esmo.org. Marginal zone lymphomas: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2020;31(1):17-29.
3. Alaggio R, Amador C, Anagnostopoulos I, et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: lymphoid neoplasms. *Leukemia*. 2022;36(7):1720-1748.
4. Campo E, Jaffe ES, Cook JR, et al. The International Consensus Classification of mature lymphoid neoplasms: a report from the Clinical Advisory Committee. *Blood*. 2022;140(11):1229-1253.
5. Kalpadakis C, Pangalis GA, Vassilakopoulos TP, Sachanas S, Angelopoulou MK. Treatment of splenic marginal zone lymphoma: should splenectomy be abandoned? *Leuk Lymphoma*. 2014;55(7):1463-1470.
6. Arcaini L, Rossi D, Paulli M. Splenic marginal zone lymphoma: from genetics to management. *Blood*. 2016;127(17):2072-2081.
7. Montalban C, Abraira V, Arcaini L, et al; Splenic Marginal Zone Lymphoma Study Group SMZLSG. Simplification of risk stratification for splenic marginal zone lymphoma: a point-based score for practical use. *Leuk Lymphoma*. 2014;55(4):929-931.
8. Muntañola A, Villalobos MT, González-Villambrosia S, et al; GELTAMO group Grupo Español de Linfomas y Trasplante de Médula Ósea. Low-risk HPLLs/ABC score patients with splenic marginal zone lymphoma can be safely managed without treatment: results from a prospective Spanish study. *Br J Haematol*. 2023;202(4):776-784.
9. Arribas AJ, Rinaldi A, Mensah AA, et al. DNA methylation profiling identifies two splenic marginal zone lymphoma subgroups with different clinical and genetic features. *Blood*. 2015;125(12):1922-1931.
10. Bonfiglio F, Bruscaggin A, Guidetti F, et al. Genetic and phenotypic attributes of splenic marginal zone lymphoma. *Blood*. 2022;139(5):732-747.
11. Duran-Ferrer M, Clot G, Nadeu F, et al. The proliferative history shapes the DNA methylome of B-cell tumors and predicts clinical outcome. *Nat Cancer*. 2020;1(11):1066-1081.
12. Machado HE, Mitchell E, Øbro NF, et al. Diverse mutational landscapes in human lymphocytes. *Nature*. 2022;608(7924):724-732.
13. Parry M, Rose-Zerilli MJ, Ljungström V, et al. Genetics and prognostication in splenic marginal zone lymphoma: revelations from deep sequencing. *Clin Cancer Res*. 2015;21(18):4174-4183.
14. Cawthon RM. Telomere measurement by quantitative PCR. *Nucleic Acids Res*. 2002;30(10):e47.
15. Strefford JC, Kadalayil L, Forster J, et al. Telomere length predicts progression and overall survival in chronic lymphocytic leukemia: data from the UK LRF CLL4 trial. *Leukemia*. 2015;29(12):2411-2414.

16. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.

17. Godec J, Tan Y, Liberzon A, et al. Compendium of immune signatures identifies conserved and species-specific biology in response to inflammation. *Immunity*. 2016;44(1):194-206.

18. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740.

19. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417-425.

20. Bryant D, Smith L, Rogers-Broadway KR, et al. Network analysis reveals a major role for 14q32 cluster miRNAs in determining transcriptional differences between IGHV-mutated and unmutated CLL. *Leukemia*. 2023;37(7):1454-1463.

21. Wojdacz TK, Amarasinghe HE, Kadalayil L, et al. Clinical significance of DNA methylation in chronic lymphocytic leukemia patients: results from 3 UK clinical trials. *Blood Adv*. 2019;3(16):2474-2481.

22. Hovestadt V, Zapatka M. Enhanced copy-number variation analysis using illumina DNA methylation arrays. http://bioconductor.org/packages/conumee/

23. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun*. 2015;6:8971.

24. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.

25. Robbe P, Ridout KE, Vavoulis DV; Genomics England Research Consortium, CLL pilot consortium. Whole-genome sequencing of chronic lymphocytic leukemia identifies subgroups with distinct biological and clinical features. *Nat Genet*. 2022;54(11):1675-1689.

26. Vavoulis DV, Cutts A, Taylor JC, Schuh A. A statistical approach for tracking clonal dynamics in cancer using longitudinal next-generation sequencing data. *Bioinformatics*. 2021;37(2):147-154.

27. Parry M, Rose-Zerilli MJ, Gibson J, et al. Whole exome sequencing identifies novel recurrently mutated genes in patients with splenic marginal zone lymphoma. *PLoS One*. 2013;8(12):e83244.

28. Jaramillo Oquendo C, Parker H, Oscier D, Ennis S, Gibson J, Strefford JC. Systematic review of somatic mutations in splenic marginal zone lymphoma. *Sci Rep*. 2019;9(1):10444.

29. Clipson A, Wang M, de Leval L, et al. KLF2 mutation is the most frequent somatic change in splenic marginal zone lymphoma and identifies a subset with distinct genotype. *Leukemia*. 2015;29(5):1177-1185.

30. Kiel MJ, Velusamy T, Betz BL, et al. Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma. *J Exp Med*. 2012;209(9):1553-1565.

31. Salido M, Baró C, Oscier D, Stamatopoulos K, Dierlamm J, Matutes E, et al. Cytogenetic aberrations and their prognostic value in a series of 330 splenic marginal zone B-cell lymphomas: a multicenter study of the Splenic B-Cell Lymphoma Group. *Blood*. 2010;116(9):1479-1488.

32. Watkins AJ, Huang Y, Ye H, Chanudet E, Johnson N, Hamoudi R, et al. Splenic marginal zone lymphoma: characterization of 7q deletion and its value in diagnosis. *J Pathol*. 2010;220(4):461-474.

33. Rossi D, Trifonov V, Fangazio M, et al. The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. *J Exp Med*. 2012;209(9):1537-1551.

34. Beà S, Valdés-Mas R, Navarro A, et al. Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. *Proc Natl Acad Sci U S A*. 2013;110(45):18250-18255.

35. Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2011;475(7354):101-105.

36. Tiacci E, Trifonov V, Schiavoni G, et al. BRAF mutations in hairy-cell leukemia. *N Engl J Med*. 2011;364(24):2305-2315.

37. Morin RD, Mendez-Lago M, Mungall AJ, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature*. 2011;476(7360):298-303.

38. Chapman MA, Lawrence MS, Keats JJ, et al. Initial genome sequencing and analysis of multiple myeloma. *Nature*. 2011;471(7339):467-472.

39. Okosun J, Wolfson RL, Wang J, et al. Recurrent mTORC1-activating RRAGC mutations in follicular lymphoma. *Nat Genet*. 2016;48(2):183-188.

40. Ying ZX, Jin M, Peterson LF, et al. Recurrent mutations in the MTOR regulator RRAGC in follicular lymphoma. *Clin Cancer Res*. 2016;22(21):5383-5393.

41. Ortega-Molina A, Deleyto-Seldas N, Carreras J, et al. Oncogenic Rag GTPase signaling enhances B cell activation and drives follicular lymphoma sensitive to pharmacological inhibition of mTOR. *Nat Metab*. 2019;1(8):775-789.

42. Ortega-Molina A, Lebrero-Fernández C, Sanz A, et al. Inhibition of Rag GTPase signaling in mice suppresses B cell responses and lymphomagenesis with minimal detrimental trade-offs. *Cell Rep*. 2021;36(2):109372.

43. Haider Z, Wästerlid T, Spångberg LD, et al. Whole-genome informed circulating tumor DNA analysis by multiplex digital PCR for disease monitoring in B-cell lymphomas: a proof-of-concept study. *Front Oncol*. 2023;13:1176698.

44. Nakajima T, Maekita T, Oda I, et al. Higher methylation levels in gastric mucosae significantly correlate with higher risk of gastric cancers. *Cancer Epidemiol Biomarkers Prev*. 2006;15(11):2317-2321.

45. Kim YT, Park JY, Jeon YK, et al. Aberrant promoter CpG island hypermethylation of the adenomatosis polyposis coli gene can serve as a good prognostic factor by affecting lymph node metastasis in squamous cell carcinoma of the esophagus. *Dis Esophagus*. 2009;22(2):143-150.

46. Tanabe K, Shinsato Y, Furukawa T, et al. Filamin C promotes lymphatic invasion and lymphatic metastasis and increases cell motility by regulating Rho GTPase in esophageal squamous cell carcinoma. *Oncotarget*. 2017;8(4):6353-6363.

47. Qi Y, Xu F, Chen L, et al. Quantitative proteomics reveals FLNC as a potential progression marker for the development of hepatocellular carcinoma. *Oncotarget*. 2016;7(42):68242-68252.

48. Vanaja DK, Ehrich M, Van den Boom D, et al. Hypermethylation of genes for diagnosis and risk stratification of prostate cancer. *Cancer Invest*. 2009; 27(5):549-560.

49. Ai J, Jin T, Yang L, et al. Vinculin and filamin-C are two potential prognostic biomarkers and therapeutic targets for prostate cancer cell migration. *Oncotarget*. 2017;8(47):82430-82436.

50. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res*. 2019;47(D1):D941-D947.

51. Sermer D, Pasqualucci L, Wendel HG, Melnick A, Younes A. Emerging epigenetic-modulating therapies in lymphoma. *Nat Rev Clin Oncol*. 2019; 16(8):494-507.

52. Bödör C, Grossmann V, Popov N, et al. *EZH2* mutations are frequent and represent an early event in follicular lymphoma. *Blood*. 2013;122(18): 3165-3168.

53. Puente XS, Beà S, Valdés-Mas R, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015;526(7574):519-524.

54. Gaidano G, Capello D, Gloghini A, et al. BCL-6 in aids-related lymphomas: pathogenetic and histogenetic implications. *Leuk Lymphoma*. 1998; 31(1-2):39-46.

55. Migliazza A, Lombardi L, Rocchi M, et al. Heterogeneous chromosomal aberrations generate 3' truncations of the NFKB2/lyt-10 gene in lymphoid malignancies. *Blood*. 1994;84(11):3850-3860.

56. Pirosa MC, Salehi M, Bruscaggin A, et al. Non-coding mutations of BCL6 in classical Hodgkin lymphoma. *Blood*. 2024;144(suppl 1):2960.

57. Leeman-Neill RJ, Song D, Bizarro J, et al. Noncoding mutations cause super-enhancer retargeting resulting in protein synthesis dysregulation during B cell lymphoma progression. *Nat Genet*. 2023;55(12):2160-2174.

58. Pasqualucci L, Neumeister P, Goossens T, et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature*. 2001; 412(6844):341-346.

59. Rossi D, Berra E, Cerri M, et al. Aberrant somatic hypermutation in transformation of follicular lymphoma and chronic lymphocytic leukemia to diffuse large B-cell lymphoma. *Haematologica*. 2006;91(10):1405-1409.

60. Liso A, Capello D, Marafioti T, et al. Aberrant somatic hypermutation in tumor cells of nodular-lymphocyte-predominant and classic Hodgkin lymphoma. *Blood*. 2006;108(3):1013-1020.

61. Khodabakhshi AH, Morin RD, Fejes AP, et al. Recurrent targets of aberrant somatic hypermutation in lymphoma. *Oncotarget*. 2012;3(11):1308-1319.

62. Hu BT, Lee SC, Marin E, Ryan DH, Insel RA. Telomerase is up-regulated in human germinal center B cells in vivo and can be re-expressed in memory B cells activated in vitro. *J Immunol*. 1997;159(3):1068-1071.

63. Arribas AJ, Gómez-Abad C, Sánchez-Beato M, et al. Splenic marginal zone lymphoma: comprehensive analysis of gene expression and miRNA profiling. *Mod Pathol*. 2013;26(7):889-901.

64. Zaragoza-Infante L, Agathangelidis A, Iatrou A, et al. Antigen selection reflected in the subclonal architecture of the B-cell receptor immunoglobulin gene repertoire in splenic marginal zone lymphoma. *Hemasphere*. 2025;9(5):e70147.