blood advances

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
bloodadvances@hematology.org

Integrated multi-omic profiling reveals two distinct splenic marginal zone lymphoma subgroups with prognostic relevance

Tracking no: ADV-2025-019336R2

Helen Parker (University of Southampton, United Kingdom) Amatta Mirandari (University College London, United Kingdom) Ben Stevens (University of Southampton, United Kingdom) Carolina Jaramillo-Oquendo (University of Southampton, United Kingdom) Martí Duran-Ferrer (Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Spain) Lara Buermann (University of Southampton, United Kingdom) Harindra Amarasinghe (University of Oxford, United Kingdom) Jaya Thomas (University of Southampton, United Kingdom) Louise Carr (University of Southampton, United Kingdom) Shama Syeda (University of Southampton, United Kingdom) Methusha Sakthipakan (University of Southampton, United Kingdom) Marina Parry (University of Southampton, United Kingdom) Matthew Rose-Zerilli (University of Southampton, United Kingdom) Zadie Davis (Royal Bournemouth Hospital, United Kingdom) Neil McIver-Brown (Royal Bournemouth Hospital, United Kingdom) Aliki Xochelli (CERTH | INAB, Greece) Sarah Ennis (University of Southampton, United Kingdom) Lydia Scarfò (Università Vita Salute San Raffaele, Italy) Paolo Ghia (Università Vita-Salute San Raffaele, Italy) Christina Kalpadakis (UNIVERSITY HOSPITAL OF CRETE, Greece) Gerassimos Pangalis (Athens Medical Center, Psychikon Branch, Athens, Greece., Greece) Davide Rossi (Institute of Oncology Research, Switzerland) Gianluca Gaidano (University of Eastern Piedmont, Italy) Simon Wagner (University of Leicester, United Kingdom) Matthew Ahearne (University of Leicester, United Kingdom) Marc Seifert (Essen University Hospital, Germany) Christoph Plass (German Cancer Research Center, Germany) Dieter Weichenhan (German Cancer Research Center,) Eva Kimby (Karolinska Institute, Sweden) Lesley-Ann Sutton (Karolinska Institutet, Sweden) Richard Rosenquist (Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden, Sweden) Rose-Marie Amini (Uppsala University, Sweden) Viktor Ljungström (Uppsala University, Sweden) Guy Pratt (Univeristy of Birmingham, United Kingdom) Francesco Forconi (University of Southampton, United Kingdom) Kostas Stamatopoulos (Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece., Greece) Marta Salido (Hospital del Mar, Spain) Ana Ferrer (Hospital Del Mar, Spain) Catherine Thieblemont (AP-HP, Hôpital Saint-Louis, Hemato-oncologie, DMU DHI,F-75010 Paris, France, France) Laura Hilton (BC Cancer, Canada) Ryan Morin (Simon Fraser University, Canada) Renata Walewska (Cancer Care, University Hospitals Dorset, United Kingdom) Jose Martin-Subero (IDIBAPS, Spain) David Oscier (Bioinformatics Research Facility (Bio-R), United Kingdom) Christopher Oakes (The Ohio State University, United States) Jane Gibson (Bioinformatics Research Facility (Bio-R), United Kingdom) Dean Bryant (University of Southampton, United Kingdom) Jonathan C. Strefford (University of Southampton, United Kingdom)

Abstract:
Splenic marginal zone lymphoma (SMZL) is a rare B-cell malignancy with notable genetic, epigenetic, and clinical heterogeneity. This study utilized coding and non-coding sequencing (n=74), including whole-genome sequencing (WGS) on 24 paired tumour-normal samples, targeted sequencing (n=55) and DNA methylation in 126 cases 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 non-coding mutational hotspots in BCL6, PAX5 and BACH2, linked to aberrant somatic hypermutation. At least one non-coding hotspot was detected in 69% of cases. Copy-number aberrations (CNAs) were present in 73% of cases, including del(7q) (27%), gain(3q) (17%) and trisomy 12 (13%). DNA methylation profiling revealed two epigenetic subgroups: SMZL-HR (high-risk, n=67) and SMZL-LR (low-risk, n=59). SMZL-HR was associated with adverse features such as female sex, IGHV1-2*04 usage, KLF2 mutations, del(7q), shorter telomeres, and elevated epiCMIT scores. Transcriptomic analysis highlighted enhanced cell proliferation in SMZL-HR, with enrichment of E2F and G2M checkpoint pathways and epigenetic regulation via EZH2. SMZL-HR patients had significantly shorter time to first treatment (TTFT) (HR: 1.9, p=.003) and reduced overall survival (HR: 2.5, p=.039): 85% of SMZL-HR patients 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 (HR: 2.4, p=.001). These findings demonstrate the role of DNA methylation and molecular profiling in SMZL risk stratification.

Conflict of interest: COI declared - see note

COI notes: R.R. had received honoraria from AbbVie, AstraZeneca, Illumina, Janssen and Roche. L.S. had received consultancy from AbbVie, AstraZeneca, BeiGene, Janssen and Lilly. P.G had received honoraria , consultancy and research funding from AbbVie, AstraZeneca, BeiGene, BMS, Lilly, Janssen, MSD and Roche. D.R. had received honoraria and research funding from AbbVie, AstraZeneca, BeiGene, BMS, Gilead, Janssen, Lilly and Kyte. M.A had received research funding from Pfizer. F.F had received honoraria from AbbVie, AstraZeneca, BeiGene and Janssen. R.W had received honoraria from AbbVie, AstraZeneca, BeiGene, Janssen and Secura bio. C.T. had received honoraria and/or consultancy from BMS, Janssen, Novartis, Roche, Abbvie, Gilead, Hospira, Amgen, Cellectis, Kyte, Takeda, Incyte, Bayer. J.S declares the following competing interests: Research funding: Roche Holding AG and Cambridge Epigenetix.

Preprint server: No;

Author contributions and disclosures: HP, DB, DO, & JS designed the study. HP, AM, CJO, MRZ, HEA, LC, SS, MS, MP, ZD, NMB & DW performed laboratory work. HP, DB, AM, LB, JT, CJO, VL, CO & BS performed data analysis. HP, DB, BS, AM & DO & JS wrote the manuscript. DO, FF, RW, KS, AX, LS, PG, CT, CK, GP, EK, RR, GP, DR, RMA, SW, MA, AF & MS provided study material and clinical data. DO, RW, JG, FS, LH, RM, CP, SE, CO, MDF, JIMS provided expert advice, and all authors reviewed the manuscript.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: Methylation array data, expression array data and targeted sequencing data are available at Biostudies under accession numbers E-MTAB-15791, E-MTAB-13778 and E-MTAB-16810 and E-MTAB-16840

Clinical trial registration information (if any):

Integrated multi-omic profiling reveals two distinct splenic marginal zone lymphoma subgroups with prognostic relevance

Helen Parker[1*], Amatta Mirandari[1*], Benjamin Stevens[1*], Carolina Jaramillo Oquendo[2], Marti 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 Scarfo[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], Francesco Forconi[1], Kostas Stamatopoulos[6], Marta Salido[22], Ana Ferrer[22], Catherine Thieblemont[23], Laura Hilton[24], Ryan Morin[24], Renata Walewska[5], José Ignacio Martín-Subero[3, 25], David Oscier[5], Christopher Oakes[26], Jane Gibson[1, 27#], Dean Bryant[1#] and Jonathan C Strefford[1, 27#].

[1] Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, UK, [2] Genomic Informatics, Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK, [3] Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain, [4] Radcliffe Department of Investigative Medicine, University of Oxford, Oxford, UK, [5] Department of Molecular Pathology, University Hospitals Dorset, Bournemouth, UK, [6] Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece, [7] Unit of B Cell Neoplasia and Strategic Research Program on CLL, Comprehensive Cancer Center, IRCCS Ospedale San Raffaele, Milano, Italy, [8] Medical School, Università Vita-Salute San Raffaele, Milano, Italy, [9] University Hospital, University of Crete, Heraklion, Greece,[10] Department of Haematology, Athens Medical Center, Athens, Greece, [11] Università della Svizzera Italiana, Institute of Oncology, Bellinzona, Switzerland, [12] Amedeo Avogadro University of Eastern Piedmont, Novara, Italy,[13] Leicester Cancer Research Centre, Leicester Royal Infirmary, Leicester, LE1 5WW, UK, [14] Institut fur zellbiologie, Universitatsklinikum Essen, Germany, [15] Epigenomics and Cancer Risk Factors, German Cancer Research Center (DKFZ), Heidelberg, Germany, [16] Unit of Hematology, Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden, [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; Department of Haematology, Heart of England NHS Foundation Trust, Birmingham, United Kingdom, [22] Laboratori de Citologia Hematològica I Citogenètica Molecular, Servei de Patologia, Hospital del Mar; GRETNHE, IMIM, Barcelona, Spain, [23] Université Paris Cité, Paris, France; Assistante Publique – Hôpitaux de Paris, Hôpital Saint-Louis, Service d'hémato-oncologie,. Paris, France, [24] Simon Fraser University, Burnaby, British Columbia, Canada, *[25] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain, [26] Division of Hematology, Department of Internal Medicine, The Ohio State University, Columbus, OH, USA.* [27] Bioinformatics Research Facility (Bio-R), Faculty of Medicine, University of Southampton, UK

*Denotes shared first authorship, [#] Denotes shared senior authorship

Corresponding Author: Professor Jonathan C Strefford, Cancer Genomics Group, MP824 Somers Building, Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, UK. JCS@soton.ac.uk T: +44 (0)23 8120 5246 | F: +44 (0)23 8079 5152

Abstract; 240 words

Text; 4375 words

Figures/tables; 5/1

References; 62

Scientific category; Lymphoid Neoplasia

For original data, please contact jcs@soton.ac.uk

Running title; Multi-Omic Profiling Reveals Two SMZL Subgroups

Data Sharing

Methylation array data are available at Biostudies under accession number E-MTAB-15791. All other data is available on request.

Key Points

- We identify two robust SMZL subgroups, with distinct methylation, mutational, chromosomal, immunogenetic and transcriptional landscapes

- The high-risk subgroup has poorer survival, higher transformation risk, and longer treatment time, surpassing other molecular models

Abstract

Splenic marginal zone lymphoma (SMZL) is a rare B-cell malignancy with notable genetic, epigenetic, and clinical heterogeneity. This study utilized coding and non-coding sequencing (n=74), including whole-genome sequencing (WGS) on 24 paired tumour-normal samples, targeted sequencing (n=55) and DNA methylation in 126 cases 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 non-coding mutational hotspots in *BCL6*, *PAX5* and *BACH2*, linked to aberrant somatic hypermutation. At least one non-coding hotspot was detected in 69% of cases. Copy-number aberrations (CNAs) were present in 73% of cases, including del(7q) (27%), gain(3q) (17%) and trisomy 12 (13%). DNA methylation profiling revealed two epigenetic subgroups: SMZL-HR (high-risk, n=67) and SMZL-LR (low-risk, n=59). SMZL-HR was associated with adverse features such as female sex, IGHV1-2*04 usage, *KLF2* mutations, del(7q), shorter telomeres, and elevated epiCMIT scores. Transcriptomic analysis highlighted enhanced cell proliferation in SMZL-HR, with enrichment of E2F and G2M checkpoint pathways and epigenetic regulation via *EZH2*. SMZL-HR patients had significantly shorter time to first treatment (TTFT) (HR: 1.9, p=.003) and reduced overall survival (HR: 2.5, p=.039): 85% of SMZL-HR patients 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 (HR: 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 10–15 years although only 40–50% of deaths are disease or treatment-related. However approximately 30% of cases have a shorter survival despite treatment and the prognosis is especially poor for the 5–15% of cases that undergo transformation to a large cell lymphoma (5, 6).

Scoring systems based on routinely measured parameters namely hemoglobin (Hb), lactate dehydrogenase (LDH) and Albumin (6), and Hb, platelet count, LDH and extrahilar lymphadenopathy (7) can divide patients into risk groups with differing 5-year cause-specific or lymphoma specific survivals. 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 watch and wait (8). Consistent with this 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, prognosis that may have potential value in clinical trial design and as therapeutic targets. Their prognostic significance is exemplified by two seminal studies. Arribas et al (9) identified two clusters of patients with differing levels of promotor methylation, showing that higher promotor methylation was associated with IGHV1-02 usage, *NOTCH2* mutations, 7q31-32 loss, histologic transformation and shorter overall survival. Bonfiglio et al (10) described two major genetic clusters and two differing splenic immune microenvironments. Those cases whose mutations involved *NF-kB* genes or *KLF2* and had an immune suppressive microenvironment had a significantly higher 10-year mortality than a matched general population.

We now seek to extend these findings by using the first whole genome sequencing analysis of tumour-germ-line duos in SMZL and employing more recent DNA methylation arrays with genome-wide coverage; the latter of which permits the analysis of EpiCMIT (epigenetically determined Cumulative MIToses), a DNA methylation-based mitotic clock which reflects the proliferative history of B-cell tumour samples (11). Our study includes untreated patients and integrates genetic, immunogenetic, phenotypic, transcriptomic, and epigenetic data to delineate more precise molecular subgroupings and their clinical implications.

Methods

Patient Cohort

Our cohort includes 133 SMZL patients, diagnosed between 1992 - 2021 at centers with expertise in the disease, meeting established diagnostic criteria. **(Table 1, Fig 1A)**. Informed consent was obtained from all patients in accordance with the Helsinki declaration, and the study was approved by regional and local ethics committees of participating institutions. Patients were sampled pre-treatment. Prior to DNA extraction (DNeasy blood and tissue kit, Qiagen, Hilden, Germany), the SMZL cells were purified from the peripheral blood (n=116) or spleen (n=17) using the EasySep Human B Cell enrichment kit without CD43 depletion (StemCell Technologies, Cambridge, UK). Tumor purity > 85% was confirmed in all cases by flow cytometry (CD19+/CD45+) (**Fig S1**).

Whole-Genome Sequencing

Whole-genome sequencing (WGS) was performed on 24 unselected SMZL cases (termed '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 at Uppsala University, Sweden). Data was analysed using custom pipelines based on GATK Best Practices Workflows for somatic variant calling (**Supp Methods**). The coding tumor mutational burden (TMB) per Mb and Kataegis regions were identified using MAFTools v2.10.0. Mutational signatures were analyzed using SigProfilerExtractor (v.1.1.4) and SigMiner (v.2.1.3) and compared to COSMIC (12). Copy-number alterations (CNA) were called from WGS data using Battenberg (v2.2.10) and PhyloWGS (v1.0-rc1) was used to infer phylogenetic trees and determine the sub-clonal composition and evolution of individual samples **(Supp Methods)**. Structural variants (SVs) were called using Delly (v.1.0.3). Each SVs required at least five reads covering each breakpoint. Only balanced rearrangements or those resulting in a concomitant CNA were considered. Each SV was manually curated by two experts.

Targeted re-sequencing

An additional 105 patients were screened using two Haloplex re-sequencing panels (Agilent Technologies, Santa Clara, CA, USA). Both panels targeted 45 genes/genomic loci selected for their established relevance in SMZL and other mature B-cell tumours, as previously described (13). The second panel also captured 15,413 novel coding and non-coding genes/genomic loci based on our WGS data and was applied to 50 patients (termed 'WGS extended cohort'). Mutations were identified using custom pipelines (**Supp Methods**). Patients were assigned a mutation-based classification, termed NNK-like, based upon the findings of Bonfiglio and colleagues (**Supp Methods**) (10).

Telomere length analysis

Telomere length (TL) was quantified in 108 cases using monochrome multiplex PCR (MMQ-PCR) (**Supp Methods**) (14). 82 CLL cases with Single Telomere Length Analysis (STELA) data were used for absolute TL extrapolation (15). Median TL (3.08kb) across the TL cohort was applied as the long/short TL cutoff.

Transcriptomic analysis

PolyA enriched mRNA sequencing libraries were prepared from 15 patients using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, Massachusetts, USA) before sequencing on an Illumina NovaSeq6000. Data was aligned and counted using STAR aligner v2.7.10b. Downstream analysis was performed in Rv4.3.1 (RRID:SCR_001905) using the EdgeR and GSEA packages (**Supp Methods**) (16-19).

DNA methylation profiling

DNA from 126 samples were processed using the Infinium Human Methylation 450 BeadChip (n=95) or Infinium Methylation EPIC BeadChip (n=31) (Illumina, Hayward, CA, USA) platforms as previously reported (20, 21). Data processing was performed using RnBeads v2.93 (RRID:SCR_010958)(22). Tumor purity was estimated using LUMP (23) and was highly consistent with FACS-based purity scores (p<0.001) (**Fig S2**). CNAs were identified using Conumee (22). Clusters were identified by hierarchical clustering of the M-values from the 10,000 most variable CpG sites (24). This was selected as the optimal approach after comparison with alternative methods (**Supp Methods**). EpiCMIT scores were calculated as per Duran-Ferrer *et al* (11). Cases were assigned to the Arribas-High (AH, n=29) or -Low (AL, n=97) methylation subgroup as per Arribas et al (9) (**Supp Methods**). For cases with methylation, CNA and mutational data available (n=122), the clonal cell fraction (CCF) was calculated for all somatic variants, considering variant allele frequency (VAF), local CNA and purity information (25, 26).

Statistics

Relationships between common clinico-biological variables (defined as ≥2% of cases (n=5)) were compared using the Fisher's exact, Mann-Whitney U test or Pearson's correlation coefficient (significance; p=0.05). Univariate cox regression analysis of the differences in time to first treatment (TTFT) (months from date of diagnosis to date of first treatment or last follow-up) and overall survival (OS) (months from date of diagnosis to date of death from any cause or last follow-up) was performed (significance; Benjamini Hochberg corrected p=0.05). Kaplan-Meier analysis was utilized to investigate OS and TTFT. Multivariate Cox Proportional Hazard models were generated for TTFT using stepwise backwards elimination. Analyses were performed in R v3.6.1 (RRID:SCR_001905).

Results

Whole genome sequencing identifies disease specific mutations

The overview of our cohort designs and methodological approaches is depicted in **Fig 1A.** In our cohort of 24 matched-normal WGS patients, we detected a median of 4,175 (range: 2,264-23,219) coding somatic mutations/sample. Consistent with prior 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%) (**Fig 1B**).

Non-coding mutations represent 99.3% (129,559/130,356) of detected variants in our WGS cohort, including recurrent mutations in the 5' UTRs of *BACH2* (17%) and the 3' UTRs of *DSG3* (12.5%) and *FGF13* (8%). Mutational hotspots, or kataegis regions, were identified in the introns of *BACH2, BCL6, PAX5, BTG2, CXCR4*, and *LTB* (**Fig 1B**). We identified mutational reference signatures from COSMIC in SMZL: SBS40 and SBS5 (both age-related) and SBS9 (linked to polymerase eta activity of somatic hypermutation) (**Fig 1B**). The SBS5 signature contributed most to the mutational landscape (median=41.6%). A significant correlation was observed between increasing IGHV mutational burden and an increased SBS9 signature proportion in individual samples ($p<.05$, $R=0.68$) (**Fig 1C**). We also detected significant enrichment of the somatic hypermutation signature (SBS-SHM) described by Machado and colleagues (12) and associated with activation-induced cytidine deaminase (AID) in the germinal center, within non-coding mutational hotspots of *BCL6, BACH2, PAX5*, and *KLF2* coding mutations (**Fig 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 one non-coding hotspot in 69% of cases, with recurrent mutations in intron 1 of *BCL6* (26%) and *PAX5* (33%) **(Table S1**). Additionally, two novel missense mutations were found in the *EZH2* coding region, bringing the total mutation count to five (p.I55M, p.K199N, p.A255T, p.A692V and GOF mutation p.Y646F).

Phylogenetic trees were inferred where possible for 21 cases with purity >0.7 and somatic SNVs >3000. Branching sub-clonal trajectories were most common (86%, n=18), with linear trajectories observed in only 3 cases (**Fig 1E**). Notably, the del(7q) and gain(3q) CNAs (n=12 and 5, respectively) originated from the main clonal population in 80% of cases. Similarly, mutations in *KLF2* were always found in the original clone with clonal cell fractions (CCF) of 1.0. In contrast, there was greater variability in the clonality of other coding drivers, including *TRAF3* and *KMT2D*, and within our non-coding hotspots.

Copy-number alterations and structural variations in SMZL

After CNA calling, utilizing both methylation array (n=126) and paired WGS data, we revealed at least one CNA (range=0-14) in 73% of patients, with 29 patients (23%) harboring a complex genome (>=3 CNA per genome). The most prevalent CNAs were del(7q) (27%), gain(3q) (16.6%), trisomy 12 (12.7%), and gain(12q) (9.5%) (**Fig 2A**). Minimally deleted or gained regions were identified on chromosomes 3q, 6q, 8p, 12q, 13q and 17p (**Table S2)**. All patients with deletions of chromosome 17p (n=8) had concomitant *TP53* mutations. While chromosome

7q deletion breakpoints were heterogenous in both size and location, a 1.75 Mb minimally deleted region (MDR) (128.57-130.32 Mb) was identified, that encompassed 24 coding genes, including *FLNC,* and 12 non-coding RNAs (**Fig 2B**). *FLNC* mutations were all missense (n=11, 4%), with a single case harbouring biallelic loss, with both a deletion and a missense mutation (p.P1060L) (**Fig 2C**).

Using our WGS cases (n=24), we focused on the further characterization of 33 SVs localized to regions of recurrent CNAs and the IGH locus. 20 translocations targeted chromosomes 3 (n=8), 7 (n=5), 12 (n=5) and 14 (n=2) (**Table S3**). Of the 12 WGS cases with del(7q), four were shown to be the result of a derivative chromosome. In case T32, the del(7q) was the result of a 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 case T14, we identified a 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 case T43, deletion of 7q and duplication of 3q were the result of a 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) (**Fig 2D**). Interestingly, in case T7T2, we identified a der(9)t(9;14)(p13.3;q32.33), which results in a translocation between *PAX5* intron 1 (chr9:37033006) and the IGH locus (chr14:105709625), a novel observation in SMZL. Inversions were less common (n=13) and notably, two patients harboured a 6.9 Mb chromosome 7 inversion with breakpoints occurring within 30 base pairs of each other (patient-T7T2 and T32) (**Table S3**). In both patients the inversion results in the disruption of the lncRNA LOC105375294 and the *ZYX* gene.

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

The presence of two epigenetically defined patient subgroups with differing clinico-biological 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 10000 most variable CpG sites identified two clusters in both groups. Bootstrapped survival analyses confirmed TTFT differences (**Suppl methods and Fig S3**). Subsequently, we performed hierarchical clustering on the entire cohort: SMZL-HR (high risk) (n=67, 53% cases) and SMZL-LR (low-risk) (n=59, 47% cases). We identified 6,891 differentially methylated CpGs sites, 6352 and 541 hyper- and hypomethylated CpGs, respectively, in the SMZL-HR group (**Fig 3A-B**).

Sixteen clinico-biological features were significantly enriched in SMZL-HR cases, 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)) (**Fig 3C**) and chromosomal alterations (del(7q), gain(3q), gain(12q) (all CNAs,

p<.01)) (**Fig 3D)**. Crucially, all patients with IGHV1-2*04 usage were assigned to the SMZL-HR subgroup (p<.001). The fraction of the SBS40 mutational signatures was also increased in SMZL-HR compared to SMZL-LR (p<.001) (**Fig S5)**. In contrast, SMZL-LR was significantly associated with a high IGHV mutational load (<97% germline identity) (p<.001) (**Fig 3C**) and showed a non-significant trend towards trisomy 12 (p=.06). The distribution of blood and spleen samples was similar between epitypes (*p*=0.34), suggesting that the observed biological differences are unlikely to be driven by tissue of origin.

The SMZL-HR epitype exhibits evidence of elevated cell proliferation

We next calculated the epiCMIT proliferative history score (11) for each patient (**Fig 4A)**. We observed that SMZL-HR patients exhibited significantly higher epiCMIT values compared to SMZL-LR patients (p<.001) (**Fig. 4B**). They were also significantly linked to the presence of the IGHV1-2*04 gene usage (p=.007) (**Fig. 4B**) and positively correlated with TMB (r=0.35, p<.001) (**Fig. 4C**). Telomere length (TL) data was available on 101 cases and ranged from 2.38-7.57 kb (median 3.06 kb) (**Fig 4Di**). Interestingly, we observed a significant negative correlation between epiCMIT and TL (R=-0.3, *p*=.002) supporting the association between cell proliferation and concomitant telomere attrition (**Fig. 4Dii**). Transcriptomic comparisons between SMZL-HR (n=10) and SMZL-LR (n=5) identified 399 differentially expressed genes (232 under-expressed, 167 overexpressed; FDR<.05, log fold change ≥1.5). Gene set enrichment analysis identified the enrichment of two pathways associated with elevated cell division; E2F targets (NES=1.98, p<.01) and G2M checkpoint (NES=2.07, p<.01), as well as epigenetic regulation via KAMMINGA_EZH2 targets (NES=1.91, p=.006) (**Table S4).**

Finally, we included epiCMIT in a comprehensive correlation analysis of 63 variables (**Table S5**) with significant associations shown in **Fig 4E and Table S6.** A total of 18 co-occurring associations (FDR<.05) were detected, many confirming the significant associations of SMZL-HR with previously discussed clinico-biological variables. Additionally, 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 chromosome 3 and 18. IGHV mutational loads <97% were mutually exclusive from IGHV1-2*04 usage, SMZL-HR status and *KLF2* mutations.

The clinical implications of molecular disease features including the SMZL-HR subgroup

During the follow-up period (median: 5 years; range: 0–24 years), 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 cases. Thirty-six patients died during the study. Although data on disease transformation was limited (n=17), we observed associations with high-risk features including IGHV1-2*04 usage (p=.005), del(7q) (p=.003), SMZL-HR (p=.009), *KLF2* (p=.02) and *TP53* (p=.03) mutation.

Univariate Cox regression analysis of 55 clinical and molecular features was performed to evaluate their prognostic significance for TTFT and OS (**Table S7 and S8**), with significant associations displayed in **Fig 5A**. Key findings showed that SMZL-HR status (HR: 1.9, p=.003), IGHV1-2*04 (HR: 2.1, p=.003), *NOTCH2* mutations (HR:

1.9, p=.013), female sex (HR: 1.8, p=.005) and gain of 3q (HR: 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, SPEN)* used by Bonfiglio and colleagues 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. Additionally, to compare our model to the 27K promoter array classification defined by Arribas in 2015, we assigned patients an AH (Arribas-High meth) (n=29) and AL (Arribas-Low meth) (n=97) status, using 86 of the top 100 most variable CpGs identified by the authors (11) (**Table S9**), showing that the AH group was significantly associated with shorter TTFT (HR: 1.8, p=0.02). Significant predictors of shorter OS include the SMZL-HR epitype (HR: 2.5, p=.039), age >60 years (HR: 18.7, p<.001), del(7q) (HR: 2.3, p=.02) and genomic complexity (>3CNA) (HR: 2.8, p=.0038).

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 SMZL-HR and SMZL-LR groups (**Fig 5B**). SMZL-HR patients also exhibited a higher frequency of therapeutic intervention (p<.001), disease transformation (p=.007), and increased mortality (p<.001) compared to SMZL-LR patients (**Fig 5C**). In analyses restricted to patients diagnosed since 2010, significant differences in TTFT and OS between epitypes were maintained, supporting the continued prognostic relevance of the score in the modern treatment era (**Fig S5**). SMZL-HR identified more poor risk patients (n=63) than the Arribas classifier (n=29) (9); UMAP analysis demonstrated that the Arribas classifier was able to capture only the most distinctly methylated cases from the SMZL-HR group (**Fig 5D),** with the more subtle methylation signatures missed by CpGs selected using the deprecated 27k array platform. Kaplan-Meier analysis confirmed that SMZL-HR patients experienced significantly shorter TTFT than SMZL-LR patients, regardless of their Arribas classification (**Fig 5E**). These findings represent a novel advancement of existing classifiers, where SMZL-HR identifies high risk patients 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 (MVA). The model incorporated the six variables predictive of shorter TTFT, as determined by univariate (UV) 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 (HR: 2.4, p=.001) and *NOTCH2* mutations (HR: 1.84, p=.04) remained the only significant independent variables associated with TTFT. We also developed a model to assess our SMZL-HR subgroup, in the context of the Arribas classification, epiCMIT scores, and NNK-like classifier. SMZL-HR remained the sole significant independent variable in the final model (HR: 1.96, p=.000), which was based on 115 patients with 84 events. Due to the low number of events, MVA for OS was not conducted.

Discussion

Herein, we conducted an analysis of the (epi)genomic landscape of SMZL, providing novel insights into its molecular and clinical features. Our WGS study, the first of its kind to include matched-normal duos, provides

crucial insights into the coding and non-coding mutational landscape of this lymphoma. We identified a median of 1.4 mutations per Mb of genomic DNA, higher than previous SMZL exome studies (31) and other indolent B-cell neoplasm such as MCL (32), CLL (33) and HCL (34), more akin to certain DLBCL (35) and MM (36) studies. In addition to coding mutations in known driver genes, we identified novel recurrent mutations in *FAM135B* and *RRAGC*. Whilst *RRAGC* mutations have been reported in 9-18% of FL (37, 38) where they have a role in deregulating MTOR signalling downstream of nutrient sensing and induce murine lymphomagenesis (39, 40), we found the same *RRAGC* variants in SMZL (n=3, 8%) (37, 41). Notably, 4% of cases harbour missense mutations in Filamin C (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 tumour progression (42-47). We also characterized recurrent mutations in the *EZH2* gene, including Y646F and A255T, previously seen in other haematological malignancies (48). The Y646F mutation is a known biomarker for Tazemetostat responsiveness and transformation in patients with FL (49, 50). However, the 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 sub-clonal trajectories in SMZL. While *KLF2* mutations typically arise from the main clonal population, several cases exhibited convergent *KLF2* lesions, highlighted by one clonal and one sub-clonal mutation, underscoring the critical role of this gene in SMZL lymphomagenesis.

Recurrent non-coding mutations were identified in the 5' UTRs of *BACH2* and *BCL6* and the 3' UTRs of *PAX5*, *DSG3* and *FGF13*. While 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* (51) and the *BCL6* non-coding mutations in various B-cell lymphomas(52-54). Validation in an additional cohort of 50 patients confirmed non-coding hotspot mutations in 69% of cases, underscoring their relevance in the disease. We identified three mutational signatures linked to aberrant somatic hypermutation, 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 aSHM may play a role in the acquisition of these mutations. While this is the first report in SMZL, aberrant SHM is well-documented in other lymphomas (55-59), where activation-induced cytidine deaminase (AID) misfires introduce mutations in key genes for B-cell function and transformation (59).

Hierarchical clustering of the methylation data and survival analysis revealed two distinct patient subgroups: SMZL-HR and SMZL-LR, with SMZL-HR patients 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, while 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 evidenced by elevated epiCMIT scores. We also found a positive correlation between tumour mutational burden (TMB) and epiCMIT scores, underscoring the relationship between cellular proliferation and the accumulation of somatic

mutations. Conversely, a negative correlation between telomere length and epiCMIT scores suggests that increased cellular proliferation may coincide with telomere attrition, potentially driving clonal evolution and disease progression. This study marks the first investigation of telomere length and epiCMIT in SMZL, revealing that epiCMIT serves as a more robust biomarker than telomere length, which can be influenced by transient telomerase upregulation in germinal center-derived B-cells (60). Additionally, transcriptomic analysis identified differentially expressed genes within the SMZL-HR subgroup involved in cell division pathways, further supporting the concept that SMZL-HR patients also possess the potential for future cellular division that likely drives the progressive phenotype observed in these patients.

By identifying two distinct disease entities, we revealed significant differences in their mutational, clinico-biological, and transcriptomic profiles, which impact clinical behaviour. Whilst the SMZL-HR subtype emerged as an independent predictor of TTFT, it did not reach significance for overall survival (OS), likely due to treatment heterogeneity and unrelated causes of death in elderly patients. Our SMZL-HR epitype substantially extends the important observations of Arribas and colleagues (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 non-coding 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, 61), highlighting its potential strength as an independent prognostic tool. Although our reconstruction of the Arribas classifier was necessarily limited to the 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 clinico-biological features. Additionally, we found a novel association between the SMZL-HR epitype and elevated epiCMIT levels, female sex and IGHV1-2*04 usage. Recent findings suggest that SMZL cases expressing the IGHV1-2*04 gene exhibit a distinct immunogenetic profile influenced by microenvironmental factors (62). Since our SMZL-HR epitype included all IGHV1-0*24 cases, it may represent a biologically distinct subgroup of SMZL with unique epigenetic and immunogenetic characteristics. However, thorough immunogenetic profiling of non-IGHV1-02*4 cases 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 tumour material was mainly blood derived, and therefore reflects complementary, pre-treatment disease biology, using samples that are more accessible in the modern era when splenectomy is rare for diagnostic but also therapeutic indications. Whilst our sub-groups should be viewed as translational rather than clinical, as robust external validation in independent cohorts is required, our study

significantly advances previous genomic and epigenetic analyses of SMZL (9, 10) in two principal ways. First, by combining the first tumour-normal whole genome sequencing dataset in SMZL with genome wide methylation profiling and broader multi-omic integration, we identify fundamental biological drivers of the disease, including noncoding aSHM hotspots, 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, characterised 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 personalised clinical management strategies in SMZL.

Acknowledgments

The authors gratefully acknowledge the patients studied herein, for primary tumour and germ-line material and clinical data. This work 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 to P.I. Foà Roberto, G.L. Ghia Paolo. We acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work. AM was funded by Cancer Immunology Talent Fund, University of Southampton. The authors gratefully acknowledge the Faculty of Medicine BIO-R, at the University of Southampton for their support and assistance in this work. RR is supported by grants from the Swedish Cancer Society, the Swedish Research Council, Region Stockholm, and Radiumhemmets Forskningsfonder, Stockholm. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory.

Authorship Contributions

HP, DB, DO, & JS designed the study. HP, AM, CJO, MRZ, HEA, LC, SS, MS, MP, ZD, NMB & DW performed laboratory work. HP, DB, AM, LB, JT, CJO, VL, CO & BS performed data analysis. HP, DB, BS, AM & DO & JS wrote the manuscript. DO, FF, RW, KS, AX, LS, PG, CT, CK, GP, EK, RR, GP, DR, RMA, SW, MA, AF & MS provided study material and clinical data. DO, RW, JG, FS, LH, RM, CP, SE, CO, MDF, JIMS provided expert advice, and all authors reviewed the manuscript.

Conflict-of-Interest Disclosure

R.R. had received honoraria from AbbVie, AstraZeneca, Illumina, Janssen and Roche. L.S. had received consultancy from AbbVie, AstraZeneca, BeiGene, Janssen and Lilly. P.G had received honoraria , consultancy and research funding from AbbVie, AstraZeneca, BeiGene, BMS, Lilly, Janssen, MSD and Roche. D.R. had received honoraria and research funding from AbbVie, AstraZeneca, BeiGene, BMS, Gilead, Janssen, Lilly and Kyte. M.A had received research funding from Pfizer. F.F had received honoraria from AbbVie, AstraZeneca, BeiGene and Janssen. R.W had received honoraria from AbbVie, AstraZeneca, BeiGene, Janssen and Secura bio. C.T. had received honoraria and/or consultancy from BMS, Janssen, Novartis, Roche, Abbvie, Gilead, Hospira, Amgen, Cellectis, Kyte, Takeda, Incyte, Bayer. J.S declares the following competing interests: Research funding: Roche Holding AG and Cambridge Epigenetix.

All other authors declare no competing interests.

References

1. Matutes E, Oscier D, Montalban C, Berger F, Callet-Bauchu E, Dogan A, et al. Splenic marginal zone lymphoma proposals for a revision of diagnostic, staging and therapeutic criteria. Leukemia. 2008;22(3):487-95.

2. Zucca E, Arcaini L, Buske C, Johnson PW, Ponzoni M, Raderer M, et al. 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, Attygalle AD, Araujo IBO, Berti E, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. Leukemia. 2022;36(7):1720-48.

4. Campo E, Jaffe ES, Cook JR, Quintanilla-Martinez L, Swerdlow SH, Anderson KC, et al. The International Consensus Classification of Mature Lymphoid Neoplasms: a report from the Clinical Advisory Committee. Blood. 2022;140(11):1229-53.

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):1463-70.

6. Arcaini L, Rossi D, Paulli M. Splenic marginal zone lymphoma: from genetics to management. Blood. 2016;127(17):2072-81.

7. Montalban C, Abraira V, Arcaini L, Domingo-Domenech E, Guisado-Vasco P, Iannitto E, et al. Simplification of risk stratification for splenic marginal zone lymphoma: a point-based score for practical use. Leuk Lymphoma. 2014;55(4):929-31.

8. Muntañola A, Villalobos MT, González-Villambrosia S, Rodríguez-Salazar MJ, Jiménez-Ubieto A, Bastidas-Mora G, et al. 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):776-84.

9. Arribas AJ, Rinaldi A, Mensah AA, Kwee I, Cascione L, Robles EF, et al. DNA methylation profiling identifies two splenic marginal zone lymphoma subgroups with different clinical and genetic features. Blood. 2015;125(12):1922-31.

10. Bonfiglio F, Bruscaggin A, Guidetti F, Terzi di Bergamo L, Faderl M, Spina V, et al. Genetic and phenotypic attributes of splenic marginal zone lymphoma. Blood. 2022;139(5):732-47.

11. Duran-Ferrer M, Clot G, Nadeu F, Beekman R, Baumann T, Nordlund J, et al. The proliferative history shapes the DNA methylome of B-cell tumors and predicts clinical outcome. Nat Cancer. 2020;1(11):1066-81.

12. Machado HE, Mitchell E, Øbro NF, Kübler K, Davies M, Leongamornlert D, et al. Diverse mutational landscapes in human lymphocytes. Nature. 2022;608(7924):724-32.

13. Parry M, Rose-Zerilli MJ, Ljungström V, Gibson J, Wang J, Walewska R, et al. Genetics and Prognostication in Splenic Marginal Zone Lymphoma: Revelations from Deep Sequencing. Clin Cancer Res. 2015;21(18):4174-83.

14. Cawthon RM. Telomere measurement by quantitative PCR. Nucleic Acids Res. 2002;30(10):e47.

15. Strefford JC, Kadalayil L, Forster J, Rose-Zerilli MJ, Parker A, Lin TT, et al. Telomere length predicts progression and overall survival in chronic lymphocytic leukemia: data from the UK LRF CLL4 trial. Leukemia. 2015;29(12):2411-4.

16. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, 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):15545-50.

17. Godec J, Tan Y, Liberzon A, Tamayo P, Bhattacharya S, Butte AJ, 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):1739-40.

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):417-25.

20. Bryant D, Smith L, Rogers-Broadway KR, Karydis L, Woo J, Blunt MD, 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):1454-63.

21. Wojdacz TK, Amarasinghe HE, Kadalayil L, Beattie A, Forster J, Blakemore SJ, et al. Clinical significance of DNA methylation in chronic lymphocytic leukemia patients: results from 3 UK clinical trials. Blood Adv. 2019;3(16):2474-81.

22. Hovestadt V ZM. conumee: Enhanced copy-number variation analysis using Illumina DNA methylation arrays. R package version 1.9.0, 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, Hu Y, Law CW, Shi W, 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, Dréau H, Kinnersley B, Denny N, et al. Whole-genome sequencing of chronic lymphocytic leukemia identifies subgroups with distinct biological and clinical features. Nat Genet. 2022;54(11):1675-89.

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):147-54.

27. Parry M, Rose-Zerilli MJ, Gibson J, Ennis S, Walewska R, Forster 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, Ashton-Key M, Wotherspoon A, Vassiliou G, 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):1177-85.

30. Kiel MJ, Velusamy T, Betz BL, Zhao L, Weigelin HG, Chiang MY, et al. Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma. J Exp Med. 2012;209(9):1553-65.

31. Rossi D, Trifonov V, Fangazio M, Bruscaggin A, Rasi S, Spina V, 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):1537-51.

32. Beà S, Valdés-Mas R, Navarro A, Salaverria I, Martín-Garcia D, Jares P, et al. Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. Proc Natl Acad Sci U S A. 2013;110(45):18250-5.

33. Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011;475(7354):101-5.

34. Tiacci E, Trifonov V, Schiavoni G, Holmes A, Kern W, Martelli MP, et al. BRAF mutations in hairy-cell leukemia. N Engl J Med. 2011;364(24):2305-15.

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

36. Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, et al. Initial genome sequencing and analysis of multiple myeloma. Nature. 2011;471(7339):467-72.

37. Okosun J, Wolfson RL, Wang J, Araf S, Wilkins L, Castellano BM, et al. Recurrent mTORC1-activating RRAGC mutations in follicular lymphoma. Nat Genet. 2016;48(2):183-8.

38. Ying ZX, Jin M, Peterson LF, Bernard D, Saiya-Cork K, Yildiz M, et al. Recurrent Mutations in the MTOR Regulator RRAGC in Follicular Lymphoma. Clin Cancer Res. 2016;22(21):5383-93.

39. Ortega-Molina A, Deleyto-Seldas N, Carreras J, Sanz A, Lebrero-Fernández C, Menéndez C, 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):775-89.

40. Ortega-Molina A, Lebrero-Fernández C, Sanz A, Deleyto-Seldas N, Plata-Gómez AB, Menéndez C, 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.

41. Haider Z, Wästerlid T, Spångberg LD, Rabbani L, Jylhä C, Thorvaldsdottir B, 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.

42. Nakajima T, Maekita T, Oda I, Gotoda T, Yamamoto S, Umemura S, et al. Higher methylation levels in gastric mucosae significantly correlate with higher risk of gastric cancers. Cancer Epidemiol Biomarkers Prev. 2006;15(11):2317-21.

43. Kim YT, Park JY, Jeon YK, Park SJ, Song JY, Kang CH, 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):143-50.

44. Tanabe K, Shinsato Y, Furukawa T, Kita Y, Hatanaka K, Minami K, 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):6353-63.

45. Qi Y, Xu F, Chen L, Li Y, Xu Z, Zhang Y, et al. Quantitative proteomics reveals FLNC as a potential progression marker for the development of hepatocellular carcinoma. Oncotarget. 2016;7(42):68242-52.

46. Vanaja DK, Ehrich M, Van den Boom D, Cheville JC, Karnes RJ, Tindall DJ, et al. Hypermethylation of genes for diagnosis and risk stratification of prostate cancer. Cancer Invest. 2009;27(5):549-60.

47. Ai J, Jin T, Yang L, Wei Q, Yang Y, Li H, et al. Vinculin and filamin-C are two potential prognostic biomarkers and therapeutic targets for prostate cancer cell migration. Oncotarget. 2017;8(47):82430-6.

48. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47(D1):D941-D7.

49. 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.

50. Bödör C, Grossmann V, Popov N, Okosun J, O'Riain C, Tan K, et al. *EZH2* mutations are frequent and represent an early event in follicular lymphoma. Blood. 2013;122(18):3165-8.

51. Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature. 2015;526(7574):519-24.

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

53. Migliazza A, Lombardi L, Rocchi M, Trecca D, Chang CC, Antonacci R, et al. Heterogeneous chromosomal aberrations generate 3' truncations of the NFKB2/lyt-10 gene in lymphoid malignancies. Blood. 1994;84(11):3850-60.

54. Pirosa MC, Salehi M, Bruscaggin A, Terzi Di Bergamo L, Jauk F, Forestieri G, et al. Non-Coding Mutations of BCL6 in Classical Hodgkin Lymphoma. Blood. 2024;144 (Supplement 1).

55. Leeman-Neill RJ, Song D, Bizarro J, Wacheul L, Rothschild G, Singh S, et al. Noncoding mutations cause super-enhancer retargeting resulting in protein synthesis dysregulation during B cell lymphoma progression. Nat Genet. 2023;55(12):2160-74.

56. Pasqualucci L, Neumeister P, Goossens T, Nanjangud G, Chaganti RS, Küppers R, et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature. 2001;412(6844):341-6.

57. Rossi D, Berra E, Cerri M, Deambrogi C, Barbieri C, Franceschetti S, et al. Aberrant somatic hypermutation in transformation of follicular lymphoma and chronic lymphocytic leukemia to diffuse large B-cell lymphoma. Haematologica. 2006;91(10):1405-9.

556 58. Liso A, Capello D, Marafioti T, Tiacci E, Cerri M, Distler V, et al. Aberrant somatic hypermutation in tumor
557 cells of nodular-lymphocyte-predominant and classic Hodgkin lymphoma. Blood. 2006;108(3):1013-20.

558 59. Khodabakhshi AH, Morin RD, Fejes AP, Mungall AJ, Mungall KL, Bolger-Munro M, et al. Recurrent targets
559 of aberrant somatic hypermutation in lymphoma. Oncotarget. 2012;3(11):1308-19.

560 60. Hu BT, Lee SC, Marin E, Ryan DH, Insel RA. Telomerase is up-regulated in human germinal center B cells
561 in vivo and can be re-expressed in memory B cells activated in vitro. J Immunol. 1997;159(3):1068-71.

562 61. Arribas AJ, Gómez-Abad C, Sánchez-Beato M, Martinez N, Dilisio L, Casado F, et al. Splenic marginal
563 zone lymphoma: comprehensive analysis of gene expression and miRNA profiling. Mod Pathol. 2013;26(7):889-
564 901.

565 62. Zaragoza-Infante L, Agathangelidis A, Iatrou A, Junet V, Pechlivanis N, Karypidou M, et al. Antigen
566 selection reflected in the subclonal architecture of the B-cell receptor immunoglobulin gene repertoire in splenic
567 marginal zone lymphoma. Hemasphere. 2025;9(5):e70147.

568

569 **Table 1. Key clinico-biological characteristics of the SMZL cohort.**

570

Variable	% of patients (numbers)
Female	59 (78/133)
Age at diagnosis—median in years [range]	68 [36-86]
Patients treated including splenectomy	72 (95/132)
Patients transformed to large cell lymphoma	26 (17/66)
Patients died	27 (36/131)
OS—median in months [range]	63 [3-293]
TTFT—median in months [range]	13 [0-179]
HPLL / ABC score; A	78 (54/69)
HPLL / ABC score; B/C	22 (15/69)
IGHV1-2*04 usage	25 (27/107)
IGHV 3-23 usage	15 (16/107)
IGHV 4-34 usage	6 (6/107)
IGHV 3-30 usage	9 (10/107)
IGHV germline identity; 100%	9 (8/94)
IGHV germline identity; 97-99.9%	39 (37/194)
IGHV germline identity; <97%	52 (49/94)
ARID1A mutations	5 (6/129)
FLNC mutations	5 (7/129)
KLF2 mutations	29 (38/129)
KMT2D mutations	21 (27/129)
MYD88 mutations	5 (7/129)
NOTCH2 mutations	21 (27/129)
TNFAIP3 mutations	13 (17/129)
TP53 mutations	16 (20/129)
TRAF3 mutations	15 (19/129)
Bonfiglio NNK-like	29 (78/129)
Presence of gain 3q	17 (21/126)
Presence of Tri3	6 (8/126)
Presence of del7q	27 (34/126)
Presence of gain 12q	10 (12/126)

Presence of Tri12	13 (16/126)
Complex genome (≥3 CNA)	6 (8/126)

SMZL-HR subgroup	53 (67/126)
Telomere length (Kb) - median [range]	3.08 [2.38-7.57]
AH classification	23 (29/126)
EpiCMIT- median [range]	0.745 [0.37-0.91]

571

Figure Legends

Figure 1. SMZL WGS; non-coding 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 non-coding mutations in **bold.** Horizontal bars below indicate the IGHV usage, sex, matched analyses and cytogenetic features for each patient. Bar charts show the base-pair context for mutations, and the contribution of the single base nucleotide (SBS) signatures to the mutational landscape. **C** Scatterplot showing the significant negative Pearson's 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 depicted by the black line **D.** Bar chart showing the significant enrichment (in red) of the SBSSHM signature in non-coding mutations, compared to the genome-wide mutational background. **E** Bar graph showing the percentage of cases studied with a branching (yellow) or linear (lilac) sub-clonal trajectory, with genomic profiles of an example of each trajectory type. CCF= The cancer cell fraction of the sub-clonal tumour population/node.

Figure 2. Coding landscape in the integrated cohort

A Frequency plot showing the genome-wide copy-number gains and losses. **B** Ideogram of chromosome 7q deletions. Blue bars represent the genomic location and size of individual deletions. The blue box indicates the MDR, and key genes located within this region are provided to the right. The pink and purple lines indicate previously described MDRs (127.5-130.8Mb [42] and 127.03M-130.07Mb [43], respectively). **C** *FLNC* lolliplot, showing the locations of 10 missense mutations (gray circles). ABD= Actin Binding Domain, RD= Rod Domain, DD=Dimerization Domain. **D** Circos plot displaying the structural variants originating from chromosome 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** Shows the 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 cases that are clonal (red) versus subclonal (blue), whilst the box and whisker plots below show the distribution of cancer cell fractions. Mutations with a CCF greater than 0.8, as indicated by the dotted green line, are considered clonal. *TP53* mutations that co-exist with a deletion of the second allele are indicated in red. **F** Trajectory plot of CCFs in double-hit *KLF2* mutants.

Figure 3. Clustering and the association with clinico-biological features.

A Heatmap, showing the top 10000 most variable CpG sites that were used in the clustering of SMZL samples, with each patient represented on the $\boldsymbol{x}$ axis. Samples and CpGs are clustered using a complete Euclidean distance method with Hierarchical clustering. The horizontal bars below show the cluster (SMZL1 (HR)= burgundy, SMZL2 (LR) = blue) **B** Scatter plot displaying the results from a principal component analysis of the top 10000 most variable CpG sites, that were used for clustering. **C** Frequency plot showing the copy-number gains and deletions (right- and left- hand panel, respectively) significantly associated with SMZL-HR and SMZL-

LR patients. **D** Frequency bar chart showing the features that are significantly associated with SMZL-HR (burgundy) and SMZL-LR (blue) patients. Significant enrichment of all features was determined using a Fishers Exact or chi-squared test, with significance levels indicated by * (p>0.05) and ** (p<0.005).

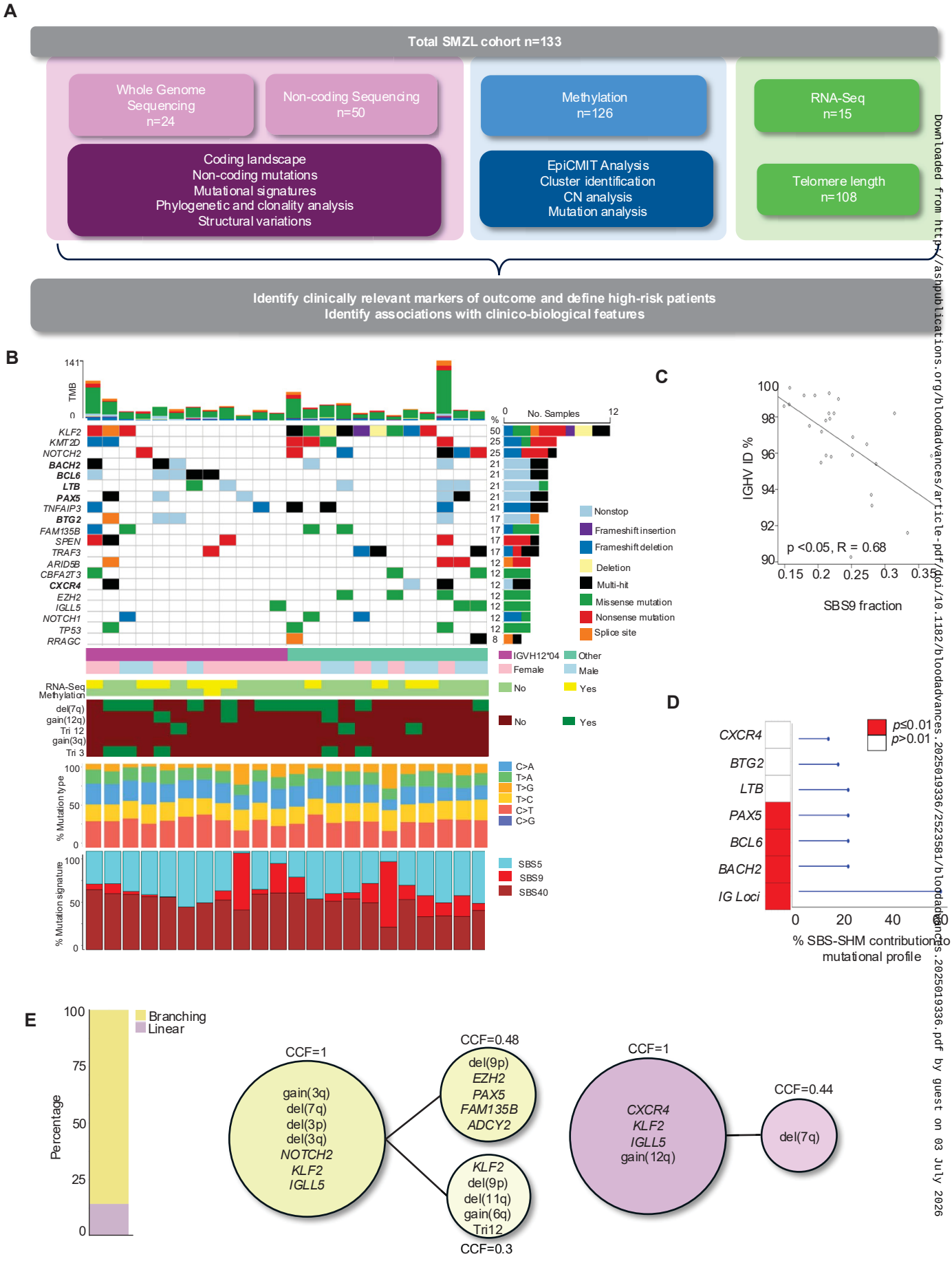
Figure 4. EpiCMIT and the association with clinico-biological features.

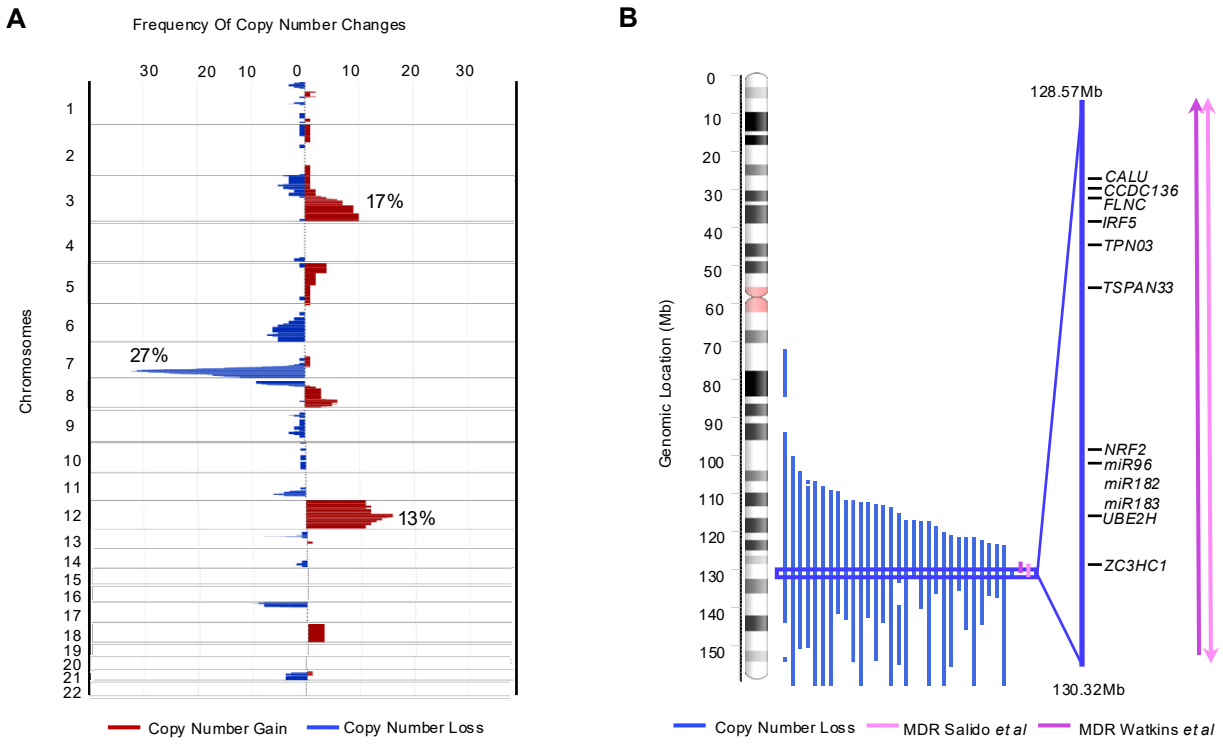
A Heatmap showing the methylation status of 174 and 1055 CpGs used to calculate the epiCMIT-hyper and -hypo scores, respectively, with each patient represented on the x axis. The horizontal bars below show the EpiCMIT score, cluster and IGHV1-2*04 status for each patient. **B** Ridgeplots showing the distribution of epiCMIT scores for variables exhibiting statistically significant differences between two status groups. The number of cases with each variable is shown. * and ** indicate p-values of p<0.05 and p<0.005, respectively. **C** Scatterplot showing the significant positive Pearson's correlation (p<0.001) between the epiCMIT score and the tumor mutational burden (TMB) per Mb in coding regions. The linear regression model is depicted by the black line, and the grey area indicates the confidence intervals. **Di** Histogram showing the frequency and distribution of telomere length (kb) in our cohort, with the median length of 3.1kb identified by the blue dotted line. **Dii** Scatterplot showing the negative Pearson's correlation (p=0.001) between high epiCMIT score and telomere length. 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) (p≤0.05, corrected p-values calculated using Benjamini-Hochberg FDR (p≤.05 *).

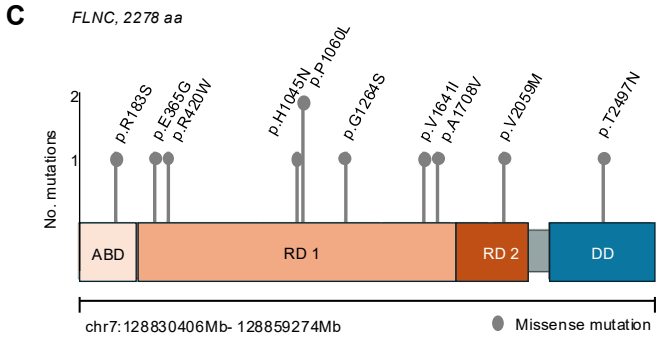
Figure 5. Clinical outcome in SMZL is associated with genomic and epigenetic features

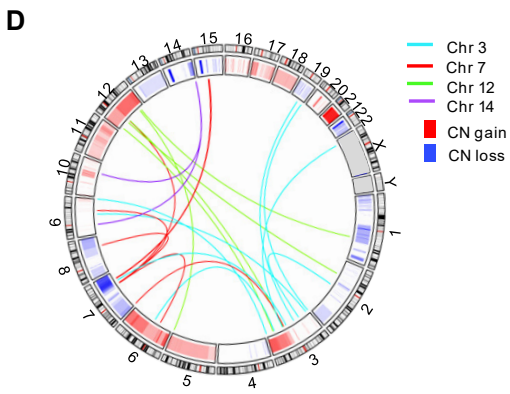
A Forest plots showing the hazard ratios for variables significant (p<0.05) for time to first treatment (TTFT) and overall survival (OS) in univariate (UV) cox regression analysis. The proportion of cases positive for each variable, and the proportion of events in those cases is shown. **B** Kaplan Meier curves for TTFT and OS (months) for SMZL-HR and SMZL-LR patients (burgundy and blue lines, respectively). **C** Stacked bar graph showing the significant difference in proportions of SMZL-HR and SMZL-LR patients that have received treatment, transformed or died. Mann-Whitney U test with Benjamini-Hochberg correction was used to determine statistical significance (p<0.05). **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 SMZL-HR patients dichotomized by Arribas high-meth (AH) and Arribas low-meth (AL) (purple and orange lines, respectively), and SMZL-LR (blue line).

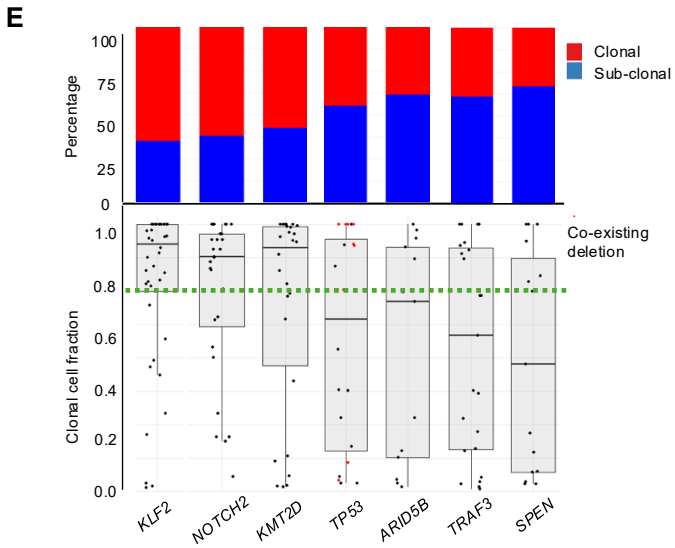
Figure 1











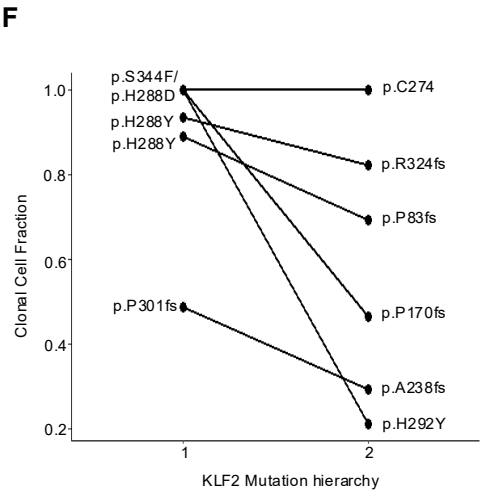
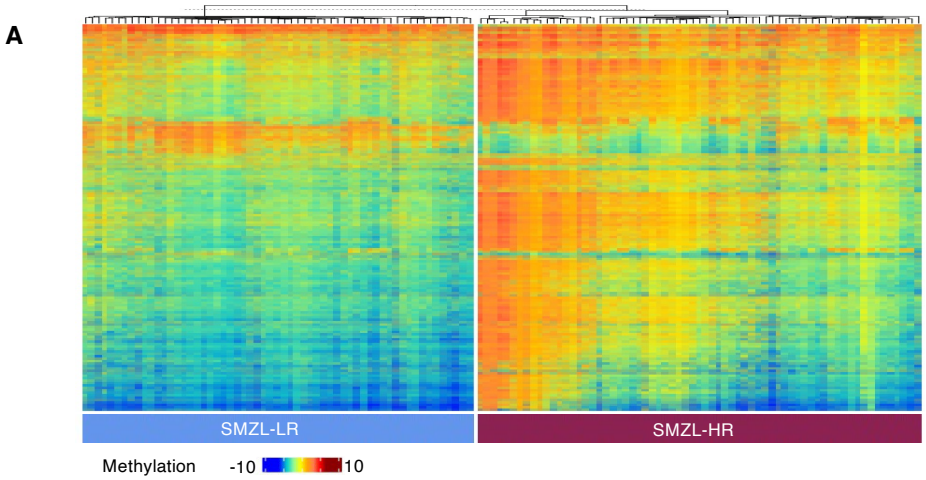
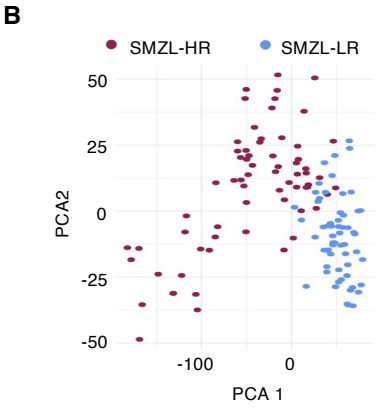
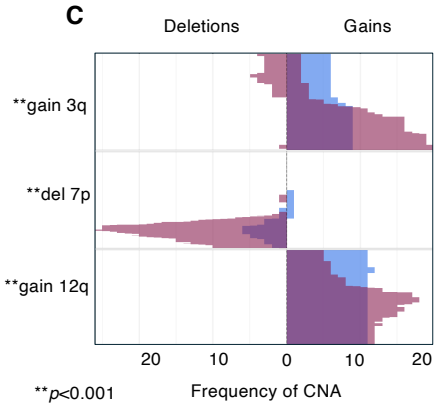
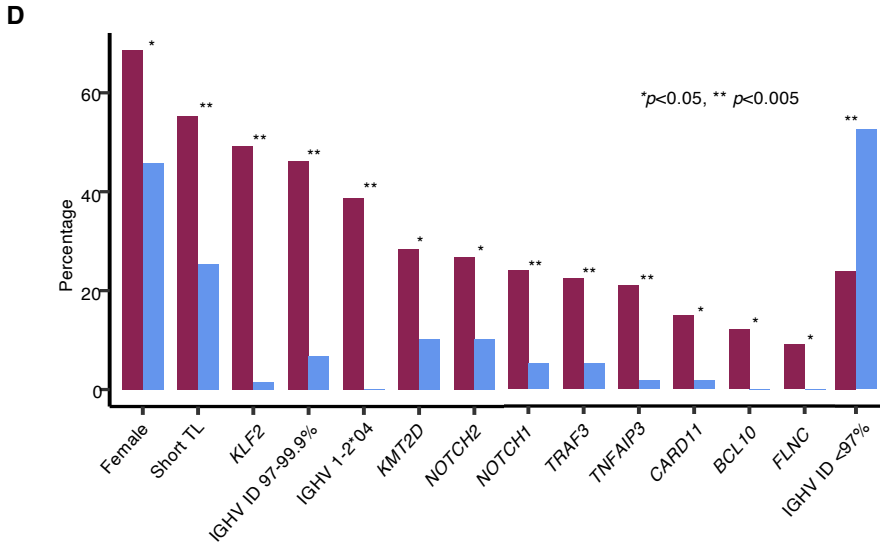


Figure 3









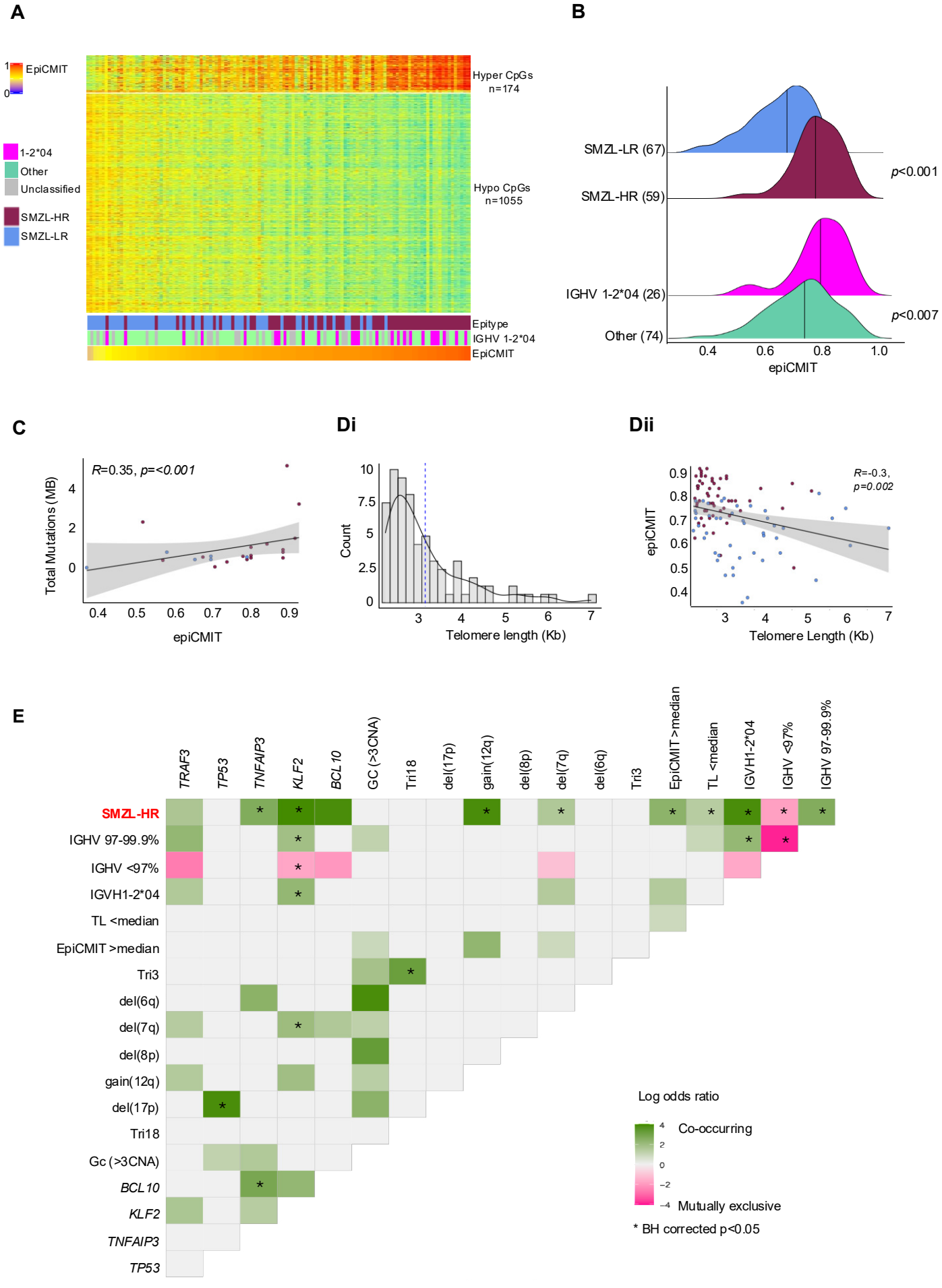


Figure 5